Saidha, Tekchand

From: Saidha, Tekchand
Sent: Monday, April 14, 2003 8:17 AM
To: Spencer, Mark
Subject: STIC copy requested - 09/622419

The stic copy for - 09/622419 is requested.

Thanks !

Tekchand Saidha
Primary Examiner
Art Unit 1652, CM1, Room No. 10D05
Mail Box 10D01
(703) 305-6595

STIC-Biotech/ChemLib

9/429

Fr m: Saidha, Tekchand
Sent: Monday, April 14, 2003 9:49 AM
To: STIC-Biotech/ChemLib
Subject: 09/622419 - sequence search request

09/622419

Please search the data base and interference files for :

SEQ ID NO : 1 & 3

Thank you !

Tekchand Saidha
Primary Examiner
Art Unit 1652, CM1, Room No. 10D05
Mail Box 10D01
(703) 305-6595

Mary Jane Ruhl
Tech. Info. Specialist, STIC
TC-1600
CM-1, Room 6A-06
Phone: 605-1155

	TYPE OF SEARCH:	VENDOR/COST (where applic.)
Searcher:_____	NA Sequences:_____	STN:_____
Phone:_____	AA Sequences:_____	DIALOG:_____
Location:_____	Structures:_____	Questel/Orbit:_____
Date Picked Up:_____	Bibliographic:_____	DRLink:_____
Date Completed: 4/22/03	Litigation:_____	Lexis/Nexis:_____
Searcher Prep/Review:_____	Full text:_____	Sequence Sys.:_____
Clerical:_____	Patent Family:_____	WWW/Internet:_____
Online time:_____	Other:_____	Other (specify):_____

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:03:55 ; Search time 3213.55 Seconds
(without alignments)
10459.998 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155
Sequence: 1 atggcaaaacacctttttac.....atgctgccggtctgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1155	100.0	1155	6	AX018960	AX018960 Sequence
2		1155	100.0	1155	6	E38344	E38344 Process for
3		1155	100.0	3720	6	AX018968	AX018968 Sequence
4		1155	100.0	4975	6	AX018972	AX018972 Sequence
5		1155	100.0	12354	1	AE000377	AE000377 Escherich
6		1155	100.0	12354	6	AX370209	AX370209 Sequence
7		1155	100.0	12354	6	AX370245	AX370245 Sequence
8		1155	100.0	12354	6	AX370258	AX370258 Sequence
9		1155	100.0	141744	1	ECU28377	U28377 Escherichia
10		1142.2	98.9	11178	1	AE005524	AE005524 Escherich
11		1142.2	98.9	266658	1	AP002563	AP002563 Escherich
12		1129.4	97.8	1152	1	ECOMETX	M98266 Escherichia
13		1043.4	90.3	1462	1	ECOMETK	K02129 E. coli met
14		929.4	80.5	22204	1	AE008842	AE008842 Salmonell
15		927.8	80.3	230050	1	AL627277	AL627277 Salmonell
16		812.6	70.4	10696	1	AE013933	AE013933 Yersinia
17		812.6	70.4	208050	1	AJ414145	AJ414145 Yersinia
18	c	763.4	66.1	10275	1	AE004133	AE004133 Vibrio ch
19		641.4	55.5	10274	1	U32797	U32797 Haemophilus
20	c	641.4	55.5	10578	1	AE006143	AE006143 Pasteurel
21		612	53.0	229426	2	AC010532	AC010532 Homo sapi
22	c	586.2	50.8	10297	1	AE004491	AE004491 Pseudomon
23	c	560	48.5	12390	1	AE014115	AE014115 Buchnera
24	c	548.8	47.5	293181	1	AP001119	AP001119 Buchnera
25	c	533	46.1	10385	1	AE011713	AE011713 Xanthomon
26	c	526.6	45.6	12556	1	AE012176	AE012176 Xanthomon
27		505	43.7	349061	1	NMA222491	AL162753 Neisseria
28	c	504.4	43.7	215050	1	AL646057	AL646057 Ralstonia
29	c	503.4	43.6	10667	1	AE002529	AE002529 Neisseria
30	c	503.4	43.6	349980	6	AX044034	AX044034 Sequence
31		489.8	42.4	9995	1	AE003891	AE003891 Xylella f
32		463.4	40.1	1149	6	AX276423	AX276423 Sequence
33		463.4	40.1	27324	6	AX276384	AX276384 Sequence
34		460	39.8	62909	6	AX067457	AX067457 Sequence
35		447.6	38.8	6040	1	RPR238756	AJ238756 Rickettsi
36		447.6	38.8	7235	1	RTY238757	AJ238757 Rickettsi
37		446	38.6	5864	1	RPFTSY	Y11784 R.prowazeki
38		446	38.6	6040	1	RPR238755	AJ238755 Rickettsi
39		446	38.6	237523	1	RPXX04	AJ235273 Rickettsi
40	c	432.4	37.4	152050	1	MLO672113	AL672113 Mesorhizo
41		422.8	36.6	346820	1	AP003008	AP003008 Mesorhizo
42		418.2	36.2	8033	1	RHE238762	AJ238762 Rickettsi
43		410.8	35.6	9007	1	RFE238763	AJ238763 Rickettsi
44	c	403.8	35.0	40356	1	SCL6	AL159139 Streptomy
45		391.4	33.9	1203	6	AX432557	AX432557 Sequence

ALIGNMENTS

RESULT 1
AX018960
LOCUS AX018960 1155 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 1 from Patent WO9942591.
ACCESSION AX018960
VERSION AX018960.1 GI:10043055
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 1155)
AUTHORS Schroeder,H.
TITLE Method for producing biotin
JOURNAL Patent: WO 9942591-A 1 26-AUG-1999;

```
            SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             1. .1155
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC07714.1"
                     /db_xref="GI:10043056"
                     /translation="MAKHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACET
                     YVKTGMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAIGKQSPD
                     INQGVDRADPLEQGAGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAEVRKNGTLPW
                     LRPDAKSQVTFQYDDGKIVGIDAVVLSTQHSEEIDQKSLQEAVMEEIIKPILPAEWLT
                     SATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTYGGMARHGGGAFSGKDPSKVDRSAA
                     YAARYVAKNIVAAGLADRCEIQVSYAIGVAEPTSIMVETFGTEKVPSEQLTLLVREFF
                     DLRPYGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAQLLRDAAGLK"
BASE COUNT      277 a    312 c    303 g    263 t
ORIGIN

  Query Match           100.0%;  Score 1155;  DB 6;  Length 1155;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-266;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    1141 GCCGGTCTGAAGTAA 1155


RESULT 2
E38344
LOCUS       E38344                  1155 bp    DNA     linear   PAT 31-JAN-2002
DEFINITION  Process for producing L-methionine by fermentation.
ACCESSION   E38344
VERSION     E38344.1  GI:18624956
KEYWORDS    JP 2000139471-A/17.
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1155)
  AUTHORS   Usuta,Y. and Kurahashi,O.
  TITLE     Process for producing L-methionine by fermentation
  JOURNAL   Patent: JP 2000139471-A 17 23-MAY-2000;
            AJINOMOTO CO INC
COMMENT     OS   Escherichia coli
            PN   JP 2000139471-A/17
            PD   23-MAY-2000
            PF   17-NOV-1998 JP 1998326717
            PR
            PI   YOSHIHIRO USUTA,OSAMU KURAHASHI
            PC   C12N15/09,C12N1/21,C12N9/04,C12N9/10,C12N9/12,C12N9/88, PC
            C12P13/12//
            PC   (C12N15/09,C12R1:19),(C12N1/21,C12R1:19),(C12P13/12,C12R1:19),
            PC   C12N15/00,
            PC   (C12N15/00,C12R1:19)
            CC
            FH   Key             Location/Qualifiers
            FT   CDS             (1). .(1152).
FEATURES             Location/Qualifiers
     source          1. .1155
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
BASE COUNT      277 a    312 c    303 g    263 t
ORIGIN

  Query Match           100.0%;  Score 1155;  DB 6;  Length 1155;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-266;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;
```

```
Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Qy   1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db   1141 GCCGGTCTGAAGTAA 1155
```

RESULT 3
AX018968
LOCUS AX018968 3720 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 9 from Patent WO9942591.
ACCESSION AX018968
VERSION AX018968.1 GI:10043063
KEYWORDS .
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE 1 (bases 1 to 3720)
AUTHORS Schroeder,H.
TITLE Method for producing biotin
JOURNAL Patent: WO 9942591-A 9 26-AUG-1999;
SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES Location/Qualifiers
source 1. .3720
/organism="Escherichia coli"
/db_xref="taxon:562"
CDS 530. .1684
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC07718.1"
/db_xref="GI:10043064"
/translation="MAKHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACET
YVKTGMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAIGKQSPD
INQGVDRADPLEQGAGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAEVRKNGTLPW
LRPDAKSQVTFQYDDGKIVGIDAVVLSTQHSEEIDQKSLQEAVMEEIIKPILPAEWLT
SATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTYGGMARHGGGAFSGKDPSKVDRSAA
YAARYVAKNIVAAGLADRCEIQVSYAIGVAEPTSIMVETFGTEKVPSEQLTLLVREFF
DLRPYGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAQLLRDAAGLK"
BASE COUNT 932 a 1019 c 954 g 815 t
ORIGIN

Query Match 100.0%; Score 1155; DB 6; Length 3720;
Best Local Similarity 100.0%; Pred. No. 1.2e-266;
Matches 1155; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    530 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 589

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    590 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 649

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    650 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 709

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    710 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 769

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    770 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 829

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    830 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 889
```

```
Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     890 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 949

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     950 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 1009

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 1069

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1070 ATCGATGCTGTCGTGCTTTCCAGTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 1129

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1130 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 1189

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1190 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 1249

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1250 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 1309

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1310 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 1369

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1370 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 1429

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1430 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 1489

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1490 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1549

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1550 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1609

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1610 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1669

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    1670 GCCGGTCTGAAGTAA 1684
```

```
RESULT 4
AX018972
LOCUS       AX018972               4975 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 13 from Patent WO9942591.
ACCESSION   AX018972
VERSION     AX018972.1  GI:10043067
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 4975)
  AUTHORS   Schroeder,H.
  TITLE     Method for producing biotin
  JOURNAL   Patent: WO 9942591-A 13 26-AUG-1999;
            SCHROEDER HARTWIG (DE); BASF AG (DE)
```

```
FEATURES             Location/Qualifiers
     source          1. .4975
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             530. .1684
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC07720.1"
                     /db_xref="GI:10043068"
                     /translation="MAKHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACET
                     YVKTGMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAIGKQSPD
                     INQGVDRADPLEQGAGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAEVRKNGTLPW
                     LRPDAKSQVTFQYDDGKIVGIDAVVLSTQHSEEIDQKSLQEAVMEEIIKPILPAEWLT
                     SATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTYGGMARHGGGAFSGKDPSKVDRSAA
                     YAARYVAKNIVAAGLADRCEIQVSYAIGVAEPTSIMVETFGTEKVPSEQLTLLVREFF
                     DLRPYGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAQLLRDAAGLK"
     CDS             1782. .2987
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC07721.1"
                     /db_xref="GI:10043069"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT     1198 a   1350 c   1329 g   1098 t
ORIGIN
```

```
  Query Match           100.0%;  Score 1155;  DB 6;  Length 4975;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-266;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     530 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 589

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     590 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 649

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     650 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 709

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     710 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 769

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     770 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 829

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     830 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 889

Qy     361 GGTCTGATGTTTGGGTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     890 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 949

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     950 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 1009

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 1069
```

```
Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1070 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 1129

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1130 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 1189

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1190 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 1249

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1250 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 1309

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1310 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 1369

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1370 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 1429

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1430 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 1489

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1490 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1549

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1550 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1609

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1610 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1669

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    1670 GCCGGTCTGAAGTAA 1684
```

```
RESULT 5
AE000377
LOCUS       AE000377               12354 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 267 of 400 of the complete
            genome.
ACCESSION   AE000377 U00096
VERSION     AE000377.1  GI:2367178
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 12354)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 12354)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 12354)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 12354)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     On Sep 9, 1997 this sequence version replaced gi:1789304.
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .12354
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     promoter        complement(39. .67)
                     /note="factor Sigma70; predicted +1 start at 3079733"
     gene            105. .989
                     /gene="yggG"
                     /note="b2936"
     CDS             105. .989
                     /gene="yggG"
                     /function="orf; Unknown"
                     /note="o294; 100 pct identical to YGGG_ECOLI SW: P25894"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAC75973.1"
                     /db_xref="GI:1789305"
                     /translation="MLACCAIYSPVRCAFPYTETERRFTCKRRIFRIILTLCGRENMK
                     IRALLVAMSVATVLTGCQNMDSNGLLSSGAEAFQAYSLSDAQVKTLSDQACQEMDSKA
                     TIAPANSEYAKRLTTIANALGNNINGQPVNYKVYMAKDVNAFAMANGCIRVYSGLMDM
                     MTDNEVEAVIGHEMGHVALGHVKKGMQVALGTNAVRVAAASAGGIVGSLSQSQLGNLG
                     EKLVNSQFSQRQEAEADDYSYDLLRQRGISPAGLATSFEKLAKLEEGRQSSMFDDHPA
                     SAERAQHIRDRMSADGIK"
     repeat_region   1008. .1154
                     /note="REP (repetitive extragenic palindromic) element;
                     contains 3 REP sequences"
     gene            complement(1195. .2115)
                     /gene="speB"
                     /note="b2937"
```

```
CDS             complement(1195. .2115)
                /gene="speB"
                /EC_number="3.5.3.11"
                /function="enzyme; Central intermediary metabolism:
                Polyamine biosynthesis"
                /note="f306; 100 pct identical to SPEB_ECOLI SW: P16936;
                CG Site No. 160"
                /codon_start=1
                /transl_table=11
                /product="agmatinase"
                /protein_id="AAC75974.1"
                /db_xref="GI:1789306"
                /translation="MSTLGHQYDNSLVSNAFGFLRLPMNFQPYDSDADWVITGVPFDM
                ATSGRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMS
                EKLQAHAEKLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGKMALVHFDAHTDTYANGCE
                FDHGTMFYTAPKEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRSVDDVIAQVKQ
                IVGDMPVYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVA
                PAYDQSEITALAAATLALEMLYIQAAKKGE"
gene            complement(2253. .4229)
                /gene="speA"
                /note="b2938"
CDS             complement(2253. .4229)
                /gene="speA"
                /EC_number="4.1.1.19"
                /function="enzyme; Central intermediary metabolism:
                Polyamine biosynthesis"
                /note="f658; 99 pct identical to SPEA_ECOLI SW: P21170; CG
                Site No. 161"
                /codon_start=1
                /transl_table=11
                /product="biosynthetic arginine decarboxylase"
                /protein_id="AAC75975.1"
                /db_xref="GI:1789307"
                /translation="MSDDMSMGLPSSAGEHGVLRSMQEVAMSSQEASKMLRTYNIAWW
                GNNYYDVNELGHISVCPDPDVPEARVDLAQLVKTREAQGQRLPALFCFPQILQHRLRS
                INAAFKRARESYGYNGDYFLVYPIKVNQHRRVIESLIHSGEPLGLEAGSKAELMAVLA
                HAGMTRSVIVCNGYKDREYIRLALIGEKMGHKVYLVIEKMSEIAIVLDEAERLNVVPR
                LGVRARLASQGSGKWQSSGGEKSKFGLAATQVLQLVETLREAGRLDSLQLLHFHLGSQ
                MANIRDIATGVRESARFYVELHKLGVNIQCFDVGGGLGVDYEGTRSQSDCSVNYGLNE
                YANNIIWAIGDACEENGLPHPTVITESGRAVTAHHTVLVSNIIGVERNEYTVPTAPAE
                DAPRALQSMWETWQEMHEPGTRRSLREWLHDSQMDLHDIHIGYSSGIFSLQERAWAEQ
                LYLSMCHEVQKQLDPQNRAHRPIIDELQERMADKMYVNFSLFQSMPDAWGIDQLFPVL
                PLEGLDQVPERRAVLLDITCDSDGAIDHYIDGDGIATTMPMPEYDPENPPMLGFFMVG
                AYQEILGNMHNLFGDTEAVDVFVFPDGSVEVELSDEGDTVADMLQYVQLDPKTLLTQF
                RDQVKKTDLDAELQQQFLEEFEAGLYGYTYLEDE"
promoter        complement(2271. .2297)
                /gene="speA"
                /note="factor Sigma70; promoter speB; documented +1 at
                3081965"
repeat_region   2759. .2791
                /note="REP (repetitive extragenic palindromic) element;
                contains 1 REP sequence"
gene            complement(4238. .4384)
                /gene="yqgB"
                /note="b2939"
CDS             complement(4238. .4384)
                /gene="yqgB"
                /function="orf; Unknown"
                /note="f48; 100 pct identical to YQGB_ECOLI SW: P46877"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75976.1"
                /db_xref="GI:1789308"
                /translation="MLNNVMKKKPVAQLERQHSLLENPCAYGLLSQFQAAIVVNCFTL
                NKII"
promoter        complement(4272. .4300)
                /gene="yqgB"
                /note="factor Sigma70; predicted +1 start at 3083966"
promoter        4339. .4368
                /note="factor Sigma70; predicted +1 start at 3084076"
promoter        complement(4469. .4496)
                /note="factor Sigma70; predicted +1 start at 3084163"
gene            4505. .4720
                /gene="yqgC"
                /note="b2940"
CDS             4505. .4720
                /gene="yqgC"
                /function="orf; Unknown"
                /note="o71; 100 pct identical to YQGC_ECOLI SW: P46878"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75977.1"
                /db_xref="GI:1789309"
                /translation="MGITSAGMQSRDAECGERIFTRTVRQVKQQTTVHYFVSPPRPPV
                KTNPQAKTLISTRLEVATRKKRKVLFI"
gene            complement(4717. .4968)
                /gene="yqgD"
                /note="b2941"
CDS             complement(4717. .4968)
                /gene="yqgD"
                /function="orf; Unknown"
                /note="f83; 100 pct identical to YQGD_ECOLI SW: P46879"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75978.1"
                /db_xref="GI:1789310"
                /translation="MVADPTTTLQVKNTGSLSVNRYGWINIWMAILGQFFTLFPLFFE
                SCLILLKTWLEIFPDNAGILRIYLLQFSAIVGYKTRRAA"
promoter        4849. .4878
```

```
Query Match             100.0%;  Score 1155;  DB 1;  Length 12354;
Best Local Similarity   100.0%;  Pred. No. 1.3e-266;
Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    .  1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5024 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 5083

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5084 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 5143

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5144 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 5203

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5204 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 5263

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5264 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 5323

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5324 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 5383

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 5443

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5444 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 5503

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5504 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 5563

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5564 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 5623
```

```
Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5624 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 5683

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5684 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5744 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 5803

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5804 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 5863

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5864 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 5923

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5924 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 5983

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5984 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 6043

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6044 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 6103

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6104 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 6163

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    6164 GCCGGTCTGAAGTAA 6178
```

```
RESULT 6
AX370209
LOCUS       AX370209               12354 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 26 from Patent WO0170776.
ACCESSION   AX370209
VERSION     AX370209.1  GI:18857386
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Levy,S.B., Barbosa,T.M. and Alekshun,M.N.
  TITLE     Nimr compositions and their methods of use
  JOURNAL   Patent: WO 0170776-A 26 27-SEP-2001;
            TRUSTEES OF TUFTS COLLEGE (US)
FEATURES             Location/Qualifiers
     source          1. .12354
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             10196. .11146
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD23853.1"
                     /db_xref="GI:18857387"
                     /translation="MIKLGIVMDPIANINIKKDSSFAMLLEAQRRGYELHYMEMGDLY
                     LINGEARAHTRTLNVKQNYEEWFSFVGEQDLPLADLDVILMRKDPPFDTEFIYATYIL
                     ERAEEKGTLIVNKPQSLRDCNEKLFTAWFSDLTPETLVTRNKAQLKAFWEKHSDIILK
                     PLDGMGGASIFRVKEGDPNLGVIAETLTEHGTRYCMAQNYLPAIKDGDKRVLVVDGEP
                     VPYCLARIPQGGETRGNLAAGGRGEPRPLTESDWKIARQIGPTLKEKGLIFVGLDIIG
```

```
                     DRLTEINVTSPTCIREIEAEFPVSITGMLMDAIEARLQQQ*
BASE COUNT     3031 a    3249 c    3190 g    2884 t
ORIGIN

  Query Match           100.0%;  Score 1155;  DB 6;  Length 12354;
  Best Local Similarity 100.0%;  Pred. No. 1.3e-266;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5024 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 5083

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5084 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 5143

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5144 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 5203

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5204 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 5263

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5264 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 5323

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5324 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 5383

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 5443

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5444 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 5503

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5504 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 5563

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5564 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 5623

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5624 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 5683

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5684 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5744 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 5803

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5804 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 5863

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5864 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 5923

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5924 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 5983

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
```

```
Db   5984 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 6043

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6044 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 6103

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6104 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 6163

Qy   1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db   6164 GCCGGTCTGAAGTAA 6178


RESULT 7
AX370245
LOCUS       AX370245               12354 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 62 from Patent WO0170776.
ACCESSION   AX370245
VERSION     AX370245.1  GI:18857412
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
  AUTHORS   Levy,S.B., Barbosa,T.M. and Alekshun,M.N.
  TITLE     Nimr compositions and their methods of use
  JOURNAL   Patent: WO 0170776-A 62 27-SEP-2001;
            TRUSTEES OF TUFTS COLLEGE (US)
FEATURES             Location/Qualifiers
     source          1. .12354
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             11183. .11818
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD23861.1"
                     /db_xref="GI:18857413"
                     /translation="MTAPFFPRILGAVAFLNQETEPLTMNLQHHFLIAMPALQDPIFR
                     RSVVYICEHNTNGAMGIIVNKPLENLKIEGILEKLKITPEPRDESIRLDKPVMLGGPL
                     AEDRGFILHTPPSNFASSIRISDNTVMTTSRDVLETLGTDKQPSDVLVALGYASWEKG
                     QLEQEILDNAWLTAPADLNILFKTPIADRWREAAKLIGVDILTMPGVAGHA"
BASE COUNT     3031 a   3249 c   3190 g   2884 t
ORIGIN

  Query Match           100.0%;  Score 1155;  DB 6;  Length 12354;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-266;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5024 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 5083

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5084 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 5143

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5144 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 5203

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5204 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 5263

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5264 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 5323

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5324 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 5383

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 5443

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5444 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 5503

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5504 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 5563

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5564 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 5623

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5624 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 5683

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5684 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5744 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 5803

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5804 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 5863

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5864 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 5923

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5924 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 5983

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5984 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 6043

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6044 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 6103

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6104 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 6163

Qy   1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db   6164 GCCGGTCTGAAGTAA 6178


RESULT 8
AX370258
LOCUS       AX370258               12354 bp    DNA     linear   PAT 16-FEB-2002
DEFINITION  Sequence 75 from Patent WO0170776.
ACCESSION   AX370258
VERSION     AX370258.1  GI:18857424
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1
```

```
  AUTHORS   Levy,S.B., Barbosa,T.M. and Alekshun,M.N.
  TITLE     Nimr compositions and their methods of use
  JOURNAL   Patent: WO 0170776-A 75 27-SEP-2001;
            TRUSTEES OF TUFTS COLLEGE (US)
FEATURES             Location/Qualifiers
     source          1. .12354
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             9425. .10183
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAD23867.1"
                     /db_xref="GI:18857425"
                     /translation="MGRRRGFLTMRIPRIYHPEPLTSHSHIALCEDAANHIGRVLRMG
                     PGQALQLFDGSNQVFDAEITSASKKSVEVKVLEGQIDDRESPLHIHLGQVMSRGEKME
                     FTIQKSIELGVSLITPLFSERCGVKLDSERLNKKLQQWQKIAIAACEQCGRNRVPEIR
                     PAMDLEAWCAEQDEGLKLNLHPRASNSINTLPLPVERVRLLIGPEGGLSADEIAMTAR
                     YQFTDILLGPRVLRTETTALTAITALQVRFGDLG"
BASE COUNT     3031 a    3249 c    3190 g    2884 t
ORIGIN

  Query Match           100.0%;  Score 1155;  DB 6;  Length 12354;
  Best Local Similarity  100.0%;  Pred. No. 1.3e-266;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5024 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 5083

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5084 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 5143

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5144 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 5203

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5204 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 5263

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5264 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 5323

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5324 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 5383

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 5443

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5444 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 5503

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5504 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 5563

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5564 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 5623

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5624 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 5683

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5684 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5744 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 5803

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGGACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5804 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 5863

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5864 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 5923

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5924 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 5983

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5984 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 6043

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6044 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 6103

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6104 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 6163

Qy    1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db    6164 GCCGGTCTGAAGTAA 6178


RESULT 9
ECU28377
LOCUS       ECU28377              141744 bp    DNA     linear   BCT 01-JUL-1995
DEFINITION  Escherichia coli K-12 genome; approximately 65 to 68 minutes.
ACCESSION   U28377
VERSION     U28377.1  GI:882431
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 87107)
  AUTHORS   Plunkett,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-JUN-1995) Guy Plunkett, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     This sequence was determined as part of the E. coli Genome Project
            (Frederick R. Blattner, director) at the University of
            Wisconsin-Madison. Supported by award HG00301 from the NIH Human
            Genome Project. The entire sequence was independently determined
            from E. coli MG1655; overlaps with other sequence determinations
            are annotated. This entry should be considered somewhat
            provisional; it will be updated and merged with others at a later
            date.
FEATURES             Location/Qualifiers
     source          1. .141744
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:562"
                     /map="approximately 65 to 68 minutes"
                     /note="This sequence comprises the following lambda
                     clones: DD656(EC18-164), DD658(EC17-126), DD660(EC30MM49),
                     DD663(EC19-62), DD665(EC27-689d), DD667(EC27-357),
                     DD670(EC30-45), DD674(EC27-1001), DD676(EC27-1160),
                     DD678(EC30K505-2), DD680(EC30K506-2), and DD682(EC17-140);
                     M13Janus vector was used for subcloning"
     misc_feature    complement(1. .5335)
                     /note="similar to GenBank Accession Number X73958
```

(ECGCVOP)"
CDS complement(<1. .292)
/note="ORF_f97p"
/codon_start=1
/transl_table=11
/protein_id="AAA69070.1"
/db_xref="GI:882432"
/translation="MAIALVTGGSRGIGRATALLLAQEGYTVAVNYQQNLHAAQEVMN
LITQAGGKAFVLQADISDENQVVAMFTAIDQHDEPLAALVNNAGILFTQCTVE"
misc_feature complement(70. .324)
/note="similar to GenBank Accession Number D21144
(ECOPYU18)"
misc_feature complement(318. .3634)
/note="similar to GenBank Accession Number L20872
(ECOGCVHP); bases 3318 to 3438 differ from the sequence
presented here, as well as from X73958 and D21144"
gene complement(559. .3432)
/gene="gcvP"
CDS complement(559. .3432)
/gene="gcvP"
/note="ORF_f957"
/codon_start=1
/transl_table=11
/protein_id="AAA69071.1"
/db_xref="GI:882433"
/translation="MTQTLSQLENSGAFIERHIGPDAAQQQEMLNAVGAQSLNALTGQ
IVPKDIQLATPPQVGAPATEYAALAELKAIASRNKRFTSYIGMGYTAVQLPPVILRNM
LENPGWYTAYTPYQPEVSQGRLEALLNFQQVTLDLTGLDMASASLLDEATAAAEAMAM
AKRVXKLKNANRFFVASDVHPQTLDVVRTRAETFGFEVIVDDAQKVLDHQDVFGVLLQ
QVGTTGEIHDYTALISELKSRKIVVSVAADIMALVLLTAPGKQGADIVFGSAQRFGVP
MGYGGPHAAFFAAKDEYKRSMPGRIIGVSKDAAGNTALRMAMQTREQHIRREKANSNI
CTSQVLLANIASLYAVYHGPVGLKRIANRIHRLTDILAAGLQQKGLKLRHAHYFDTLC
VEVADKAGVLTRAEAAEINLRSDILNAVGITLDETTTRENVMQLFNVLLGDNHGLDID
TLDKDVAHDSRSIQPAMLRDDEILTHPVFNRYHSETEMMRYMHSLERKDLALNQAMIP
LGSCTMKLNAAAEMIPITWPEFAELHPFCPPEQAEGYQQMIAQLADWLVKLTGYDAVC
MQPNSGAQGEYAGLLAIRHYHESRNEGHRDICLIPASAHGTNPASAHMAGMQVVVVAC
DKNGNIDLTDLRAKAEQAGDNLSCIMVTYPSTHGVYEETIREVCEVVHQFGGQVYLDG
ANMNAQVGITSPGFIGADVSHLNLHKTFCIPHGGGGPGMGPIGVKAHLAPFVPGHSVV
QIEGMLTRQGAVSAAPFGSASILPISWMYIRMMGAEGLKKASQVAILNANYIASRLQD
AFPVLYTGRDGRVAHECILDIRPLKEETGISELDIAKRLIDYGFHAPTMSFPVAGTLM
VEPTESESKVELDRFIDAMLAIRAEIDQVKAGVWPLEDNPLVNAPHIQSELVAEWAHP
YSREVAVFPAGVADKYWPTVKRLDDVYGDRNLFCSCVPISEYQ"
gene complement(3551. .3940)
/gene="gcvH"
CDS complement(3551. .3940)
/gene="gcvH"
/note="ORF_f129"
/codon_start=1
/transl_table=11
/protein_id="AAA69072.1"
/db_xref="GI:882434"
/translation="MSNVPAELKYSKEHEWLRKEADGTYTVGITEHAQELLGDMVFVD
LPEVGATVSAGDDCAVAESVKAASDIYAPVSGEIVAVNDALSDSPELVNSEPYAGGWI
FKIKASDESELESLLDATAYEALLEDE"
misc_feature complement(3809. .5631)
/note="similar to GenBank Accession Number M97263
(ECOTPRO)"
gene complement(3964. .5058)
/gene="gcvT"
CDS complement(3964. .5058)
/gene="gcvT"
/note="ORF_f364"
/codon_start=1
/transl_table=11
/protein_id="AAA69073.1"
/db_xref="GI:882435"
/translation="MAQQTPLYEQHTLCGARMVDFHGWMMPLHYGSQIDEHHAVRTDA
GMFDVSHMTIVDLRGSRTREFLRYLLANDVAKLTKSGKALYSGMLNASGGVIDDLIVY
YFTEDFFRLVVNSATREKDLSWITQHAEPFGIEITVRDDLSMIAVQGPNAQAKAATLF
NDAQRQAVEGMKPFFGVQAGDLFIATTGYTGEAGYEIALPNEKAADFWRALVEAGVKP
CGLGARDTLRLEAGMNLYGQEMDETISPLAANMGWTIAWEPADRDFIGREALEVQREH
GTEKLVGLVMTEKGVLRNELPVRFTDAQGNQHEGIITSGTFSPTLGYSIALARVPEGI
GETAIVQIRNREMPVKVTKPVFVRNGKAVA"

misc_feature complement(5370. .10317)
/note="similar to GenBank Accession Number D90281
(ECOPUV)"
gene complement(5506. .6708)
/gene="visC"
CDS complement(5506. .6708)
/gene="visC"
/note="ORF_f400"
/codon_start=1
/transl_table=11
/protein_id="AAA69074.1"
/db_xref="GI:882436"
/translation="MQSVDVAIVGGGMVGLAVACGLQGSGLRVAVLEQRVQEPLAANA
PPQLRVSAINAASEKLLTRLGVWQDILSRRASCYHGMEVWDKDSFGHISFDDQSMGYS
HLGHIVENSVIHYALWNKAHQSSDITLLAPAELQQVAWGENETFLTLKDGSMLTARLV
IGADGANSWLRNKADIPLTFWDYQHHALVATIRTEEPHDAVARQVFHGEGILAFLPLS
DPHLCSIVWSLSPEEAQRMQQASEDEFNRALNIAFDNRLGLCKVESARQVFPLTGRYA
RQFASHRLALVGDAAHTIHPLAGQGVNLGFMDAAELIAELKRLHRQGKDIGQYIYLRR
YERSRKHSAALMLAGMQGFRDLFSGTNPAKKLLRDIGLKLADTLPGVKPQLIRQAMGL
NDLPEWLR"
gene complement(6731. .7909)
/gene="ubiH"
CDS complement(6731. .7909)
/gene="ubiH"
/note="ORF_f392"
/codon_start=1
/transl_table=11
/protein_id="AAA69075.1"
/db_xref="GI:882437"
/translation="MSVIIVGGGMAGATLALAISRLSHGALPVHLIEATAPESHAHPG
FDGRAIALAAGTCQQLARIGVWQSLADCATAITTVHVSDRGHAGFVTLAAEDYQLAAL
GQVVELHNVGQRLFALLRKAPGVTLHCPDRVANVARTQSHVEVTLESGETLTGRVLVA
ADGTHSALATACGVDWQQEPYEQLAVIANVATSVAHEGRAFERFTQHGPLAMLPMSDG
RCSLVWCHPLERREEVLSWSDEKFCRELQSAFGWRLGKITHAGKRSAYPLALTHAARS
ITHRTVLVGNAAQTLHPIAGQGFNLGMRDVMSLAETLTQAQERGEDMGDYGVLCRYQQ
RRQSDREATIGVTDSLVHLFANRWAPLVVGRNIGLMTMELFTPARDVLAQRTLGWVAR
"
misc_feature complement(7749. .9599)
/note="similar to GenBank Accession Number D00398
(ECOAPP2)"
gene complement(7906. .9231)
/gene="pepP"
CDS complement(7906. .9231)
/gene="pepP"
/note="ORF_f441; third start codon"
/codon_start=1
/transl_table=11
/protein_id="AAA69076.1"
/db_xref="GI:882438"
/translation="MSEISRQEFQRRRQALVEQMQPGSAALIFAAPEVTRSADSEYPY
RQNSDFWYFTGFNEPEAVLVLIKSDDTHNHSVLFNRVRDLTAEIWFGRRLGQDAAPEK
LGVDRALAFSEINQQLYQLLNGLDVVYHAQGEYAYADVIVNSALEKLRKGSRQNLTAP
ATMIDWRPVVHEMRLFKSPEEIAVLRRAGEITAMAHTRAMEKCRPGMFEYHLEGETHH
EFNRHGARYPSYNTIVGSGENGCILHYTENECEMRDGDLVLIDAGCEYKGYAGDITRT
FPVNGKFTQAQREIYDIVLESLETSLRLYRPGTSILEVTGEVVRIMVSGLVKLGILKG
DVDELIAQNAHRPFFMHGLSHWLGLDVHDVGVYGQDRSRILEPGMVLTVEPGLYIAPD
AEVPEQYRGIGIRIEDDIVITETGNENLTASVVKKPEEIEALMVAARKQ"
gene complement(9257. .9841)
/gene="ygfB"
CDS complement(9257. .9841)
/gene="ygfB"
/note="ORF_f194"
/codon_start=1
/transl_table=11
/protein_id="AAA69077.1"
/db_xref="GI:882439"
/translation="MLMSIQNEMPGYNEMNQYLNQQGTGLTPAEMHGLISGMICGGND
DSSWLPLLHDLTNEGMAFGHELAQALRKMHSATSDALQDDGFLFQLYLPDGDDVSVFD
RADALAGWVNHFLLGLGVTQPKLDKVTGETGEAIDDLRNIAQLGYDEDEDQEELEMSL
EEIIEYVRVAALLCHDTFTHPQPTAPEVQKPTLH"
CDS 10003. .10332
/note="ORF_o109"
/codon_start=1

```
                     /transl_table=11
                     /protein_id="AAA69078.1"
                     /db_xref="GI:882440"
                     /translation="MSAQPVDIQIFGRSLRVNCPPDQRDALNQAADDLNQRLQDLKER
                     TRVTNTEQLVFIAALNISYELAQEKAKTRDYAASMEQRIRMLQQTIEQALLEQGRITE
                     KTNQNFE"
     misc_feature    10266. .11219
                     /note="similar to GenBank Accession Number M12965
                     (ECOSSR)"
     misc_feature    10374. .10557
                     /note="similar to GenBank Accession Number X01238
                     (ECRRN6S)"
     gene            10374. .10556
                     /gene="ssr"
     misc_RNA        10374. .10556
                     /gene="ssr"
                     /product="stable 6S RNA"

  Query Match             100.0%;  Score 1155;  DB 1;  Length 141744;
  Best Local Similarity   100.0%;  Pred. No. 1.4e-266;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41101 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 41160

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41161 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 41220

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41221 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 41280

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41281 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 41340

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41341 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 41400

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41401 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 41460

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41461 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 41520

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41521 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 41580

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41581 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 41640

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41641 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 41700

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41701 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 41760

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41761 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 41820

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41821 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 41880

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41881 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 41940

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41941 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 42000

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42001 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 42060

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42061 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 42120

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42121 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 42180

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42181 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 42240

Qy    1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db   42241 GCCGGTCTGAAGTAA 42255


RESULT 10
AE005524
LOCUS       AE005524               11178 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 143
            of 290.
ACCESSION   AE005524 AE005174
VERSION     AE005524.1  GI:12517476
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 11178)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 11178)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .11178
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            complement(188. .2179)
                     /gene="tktA"
                     /note="Z4279"
```

```
CDS             complement(188. .2179)
                /gene="tktA"
                /function="enzyme; Central intermediary metabolism:
                Non-oxidative branch, pentose pwy"
                /note="Residues 1 to 663 of 663 are 99.69 pct identical to
                residues 1 to 663 of 663 from Escherichia coli K-12 Strain
                MG1655: B2935"
                /codon_start=1
                /transl_table=11
                /product="transketolase 1 isozyme"
                /protein_id="AAG58065.1"
                /db_xref="GI:12517477"
                /translation="MSSRKELANAIRALSMDAVQKAKSGHPGAPMGMADIAEVLWRDF
                LKHNPQNPSWADRDRFVLSNGHGSMLIYSLLHLTGYDLPMEELKNFRQLHSKTPGHPE
                VGYTAGVETTTGPLGQGIANAVGMAIAEKTLAAQFNRPGHDIVDHYTYAFMGDGCMME
                GISHEVCSLAGTLKLGKLIAFYDDNGISIDGHVEGWFTDDTAMRFEAYGWHVIRDIDG
                HDAASIKRAVEEARAVTDKPSLLMCKTIIGFGSPNKAGTHDSHGAPLGDAEIALTREQ
                LGWKYAPFEIPSEIYAQWDAKEAGQAKESAWNEKFAAYAKAYPQEAAEFTRRMKGEMP
                SDFDAKAKEFIAKLQANPAKIASRKASQNAIEAFGPLLPEFLGGSADLAPSNLTLWSG
                SKAINEDAAGNYIHYGVREFGMTAIANGISLHGGFLPYTSTFLMFVEYARNAVRMAAL
                MKQRQVMVYTHDSIGLGEDGPTHQPVEQVASLRVTPNMSTWRPCDQVESAVAWKYGVE
                RQDGPTALILSRQNLAQQERTEEQLANIARGGYVLKDCAGQPELIFIATGSEVELAVA
                AYEKLTAEGVKARVVSMPSTDAFDKQDAAYRESVLPKAVTARVAVXAGIADYWYKYVG
                LNGAIIGMTTFGESAPAELLFEEFGFTVDNVVAKAKELL"
gene            2331. .3215
                /gene="yggG"
                /note="Z4280"
CDS             2331. .3215
                /gene="yggG"
                /function="orf; Unknown function"
                /note="Residues 1 to 294 of 294 are 99.31 pct identical to
                residues 1 to 294 of 294 from Escherichia coli K-12 Strain
                MG1655: B2936"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG58066.1"
                /db_xref="GI:12517478"
                /translation="MLACCAIYSPVRCAFPYTETERRFTCKRRIFRIILTLCGRKNMK
                IRALLVAMSVATVLTGCQNMDSNGLLSSGAEAFQAYSLSDAQVKTLSDQACQEMDSKA
                TIAPANSEYAKRLTTIANALGNNINGQPVNYKVYMAKDVNAFAMANGCIRVYSGLMDM
                MTDNEVEAVIGHEMGHVALGHVKKGMQVALGTNAVRVAAASAGGIVGSLSQSQLGDLG
                EKLVNSQFSQRQEAEADDYSYDLLRQRGISPAGLATSFEKLAKLEEGRQSSMFDDHPA
                SAERAQHIRDRMSADGIK"
gene            complement(3421. .4341)
                /gene="speB"
                /note="Z4281"
CDS             complement(3421. .4341)
                /gene="speB"
                /function="enzyme; Central intermediary metabolism:
                Polyamine biosynthesis"
                /note="Residues 1 to 306 of 306 are 100.00 pct identical
                to residues 1 to 306 of 306 from Escherichia coli K-12
                Strain MG1655: B2937"
                /codon_start=1
                /transl_table=11
                /product="agmatinase"
                /protein_id="AAG58067.1"
                /db_xref="GI:12517479"
                /translation="MSTLGHQYDNSLVSNAFGFLRLPMNFQPYDSDADWVITGVPFDM
                ATSGRAGGRHGPAAIRQVSTNLAWEHNRFPWNFDMRERLNVVDCGDLVYAFGDAREMS
                EKLQAHAEKLLAAGKRMLSFGGDHFVTLPLLRAHAKHFGKMALVHFDAHTDTYANGCE
                FDHGTMFYTAPKEGLIDPNHSVQIGIRTEFDKDNGFTVLDACQVNDRSVDDVIAQVKQ
                IVGDMPVYLTFDIDCLDPAFAPGTGTPVIGGLTSDRAIKLVRGLKDLNIVGMDVVEVA
                PAYDQSEITALAAATLALEMLYIQAAKKGE"
misc_feature    4419. .5353
                /note="O-island #120; Region of the EDL933 chromosome not
                homologous to E. coli K-12 MG1655"
gene            complement(4477. .5208)
                /gene="Z4282"
CDS             complement(4477. .5208)
                /gene="Z4282"
                /function="orf; Unknown function"
                /note="No significant matches"
                /codon_start=1
                /transl_table=11
                /protein_id="AAG58068.1"
                /db_xref="GI:12517480"
                /translation="MKKWKVRSALVALIVLLAGCSSNAQYNSSASGNVGTAWGGDVHS
                TVQGVSAERAWRDPAEMIVISYSTNVPSGYDRVYSIRINELEYAIRDGNFNSLPITRV
                YDSSNNEPRYIVHARVGMNYQLYVRNYSRNTNYEIVATVDGLDVLNGKQGSLNNNGYI
                VNAGDSLAIKGFRKDKHTEAAFQFANVADSYAANSAQGDVRNTGVIGFAAFELQGQAQ
                NALPPCSGQAFPADNNGYAPPPCRK"
gene            complement(5354. .7330)
                /gene="speA"
                /note="Z4283"
CDS             complement(5354. .7330)
                /gene="speA"
                /function="enzyme; Central intermediary metabolism:
                Polyamine biosynthesis"
                /note="Residues 1 to 658 of 658 are 99.84 pct identical to
                residues 1 to 658 of 658 from Escherichia coli K-12 Strain
                MG1655: B2938"
                /codon_start=1
                /transl_table=11
                /product="biosynthetic arginine decarboxylase"
                /protein_id="AAG58069.1"
                /db_xref="GI:12517481"
                /translation="MSDDMSMGLPSSAGEHGVLRSMQEVAMSSQEASKMLRTYNIAWW
                GNNYYDVNELGHISVCPDPDVPEARVDLAQLVKTREAQGQRLPALFCFPQILQHRLRS
                INAAFKRARESYGYNGDYFLVYPIKVNQHRRVIESLIHSGEPLGLEAGSKAELMAVLA
                HAGMTRSVIVCNGYKDREYIRLALIGEKMGHKVYLVIEKMSEIAIVLDEAERLNVVPR
                LGVRARLASQGSGKWQSSGGEKSKFGLAATQVLQLVETLREAGRLDSLQLLHFHLGSQ
                MANIRDIATGVRESARFYVELHKLGVNIQCFDVGGGLGVDYEGTRSQSDCSVNYGLNE
                YANNIIWAIGDACEENGLPHPTVITESGRAVTAHHTVLVSNIIGVERNEYTVPTAPAE
                DAPRALQSMWETWQEMHEPGTRRSLREWLHDSQMDLHDIHIGYSSGTFSLQERAWAEQ
                LYLSMCHEVQKQLDPQNRAHRPIIDELQERMADKMYVNFSLFQSMPDAWGIDQLFPVL
                PLEGLDQVPERRAVLLDITCDSDGAIDHYIDGDGIATTMPMPEYDPENPPMLGFFMVG
                AYQEILGNMHNLFGDTEAVDVFVFPDGSVEVELSDEGDTVADMLQYVQLDPKTLLTQF
                RDQVKKTDLDAELQQQFLEEFEAGLYGYTYLEDE"
gene            complement(7339. .7485)
                /gene="yqgB"
                /note="Z4284"
CDS             complement(7339. .7485)
                /gene="yqgB"
                /function="orf; Unknown function"
                /note="Residues 1 to 48 of 48 are 100.00 pct identical to
                residues 1 to 48 of 48 from Escherichia coli K-12 Strain
                MG1655: B2939"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG58070.1"
                /db_xref="GI:12517482"
                /translation="MLNNVMKKKPVAQLERQHSLLENPCAYGLLSQFQAAIVVNCFTL
                NKII"
gene            7606. .7821
                /gene="yqgC"
                /note="Z4285"
CDS             7606. .7821
                /gene="yqgC"
                /function="orf; Unknown function"
                /note="Residues 1 to 71 of 71 are 100.00 pct identical to
                residues 1 to 71 of 71 from Escherichia coli K-12 Strain
                MG1655: B2940"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAG58071.1"
                /db_xref="GI:12517483"
                /translation="MGITSAGMQSRDAECGERIFTRTVRQVKQQTTVHYFVSPPRPPV
                KTNPQAKTLISTRLEVATRKKRRVLFI"
gene            complement(7818. .8069)
                /gene="yqgD"
                /note="Z4286"
CDS             complement(7818. .8069)
```

```
                     /gene="yqgD"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 83 of 83 are 98.79 pct identical to
                     residues 1 to 83 of 83 from Escherichia coli K-12 Strain
                     MG1655: B2941"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG58072.1"
                     /db_xref="GI:12517484"
                     /translation="MVADPTTTLQVKNTGSLSVNRYGWINIWMAILGQFFTRFPLFFE
                     SCLILLKTWLEIFPDNAGILRIYLLQFSAIVGYKTRRAA"
     gene            8125. .9279

  Query Match           98.9%;  Score 1142.2;  DB 1;  Length 11178;
  Best Local Similarity  99.3%;  Pred. No. 1.5e-263;
  Matches 1147;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8125 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 8184

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8185 GACCAAATCTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 8244

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8245 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 8304

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8305 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 8364

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    8365 TCCGACATGGGCTTTGACGCTAACTCCTGTGCAGTTCTGAGCGCTATCGGCAAACAGTCT 8424

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8425 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 8484

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8485 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 8544

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8545 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 8604

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    8605 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTCCAGTATGACGACGGCAAAATCGTTGGT 8664

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8665 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 8724

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db    8725 GAAGCGGTAATGGAAGAGATCATCAAGCCGATTCTGCCCGCTGAATGGCTGACTTCTGCC 8784

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          |||||||||||||||||||||||||||||| ||||||||||||| ||||||||||||||
Db    8785 ACCAAATTCTTCATCAACCCGACCGGTCGTTTTGTTATCGGTGGCCCGATGGGTGACTGC 8844

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8845 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 8904

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8905 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 8964

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    8965 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 9024

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          |||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db    9025 TACGCAATCGGCGTGGCTGAACCGACTTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 9084

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db    9085 GTGCCTTCTGAACAACTGACTCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 9144

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9145 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 9204

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9205 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 9264

Qy    1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db    9265 GCCGGTCTGAAGTAA 9279


RESULT 11
AP002563
LOCUS       AP002563               266658 bp    DNA     linear   BCT 07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 14/20.
ACCESSION   AP002563 BA000007
VERSION     AP002563.1  GI:13363121
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
```

```
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 266658)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .266658
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            complement(244. .693)
                     /gene="ECs3650"
     CDS             complement(244. .693)
                     /gene="ECs3650"
                     /note="similar to B2790_ECOLI gi|1789154 percent identity
                     100 in 149 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37073.1"
                     /db_xref="GI:13363122"
                     /translation="MAEIGIFVGTMYGNSLLVAEEAEAILTAQGHKATVFEDPELSDW
                     LPYQDKYVLVVTSTTGQGDLPDSIVPLFQGIKDSLGFQPNLRYGVIALGDSSYVNFCN
                     GGKQFDALLQEQSAQRVGEMLLIDASENPEPETESNPWVEQWGTLLS"
     gene            complement(711. .1493)
                     /gene="ECs3651"
     CDS             complement(711. .1493)
                     /gene="ECs3651"
                     /note="similar to YQCB_ECOLI gi|1789155 percent identity
                     99 in 260 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37074.1"
                     /db_xref="GI:13363123"
                     /translation="MLEILYQDEWLVAVNKPSGWLVHRSWLDRDEKVVVMQTVRDQIG
                     QHVFTAHRLDRPTSGVLLMGLSSEAGRLLAQQFEQHQIQKRYHAIVRGWLMEEAVLDY
                     PLVEELDKIADKFAREDKGPQPAVTHYRGLATVEMPVATGRYPTTRYGLVELEPKTGR
                     KHQLRRHLAHLRHPIIGDSKHGDLRQNRSGAEHFGLQRLMLHASQLSLTHPFTGEPLT
                     IHAGLDDTWMQALSQFGWRGLLPENERVEFSAPSGQDGERSS"
     gene            complement(1493. .1822)
                     /gene="ECs3652"
     CDS             complement(1493. .1822)
                     /gene="ECs3652"
                     /note="similar to B2792_ECOLI gi|1789156 percent identity
                     98 in 109 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37075.1"
                     /db_xref="GI:13363124"
                     /translation="MTTHDRVRLQLQALEALLREHQHWRNDEPQPHQFNSTQPFFMDT
                     MEPLEWLQWVLIPRMHDLLNNNQPLPGAFAVAPYYEMALATDHPQRALILAELEKLDA
                     LFADDAS"
     gene            complement(1865. .2233)
                     /gene="csrB"
     misc_RNA        complement(1865. .2233)
                     /gene="csrB"
                     /product="CsrB RNA"
                     /note="inhibitor of CsrA"
     gene            complement(2444. .2989)
                     /gene="ECs3653"
     CDS             complement(2444. .2989)
                     /gene="ECs3653"
                     /note="similar to SYD_ECOLI gi|1789157 percent identity 98
                     in 181 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="Syd protein"
                     /protein_id="BAB37076.1"
                     /db_xref="GI:13363125"
                     /translation="MDDLTAQALKDFTARYCDAWHEEHKSWPLSEELYGVPSPCIIST
                     TEDAVYWQPQPFTGEQNVNAVERAFDIVIQPTIHTFYTTQFAGDMHAQFGDIKLTLLQ
                     TWSEDDFRRVQENLIGHLVTQKRLKLPPTLFIATLEEELEVISVCNLSGEVCKETLGT
                     RKRTHLAPNLAEFLNQLKPLL"
     gene            3057. .3905
                     /gene="ECs3654"
     CDS             3057. .3905
                     /gene="ECs3654"
                     /note="similar to YQCD_ECOLI gi|1789158 percent identity
                     98 in 282 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37077.1"
                     /db_xref="GI:13363126"
                     /translation="MSSYANHQALAGLTLGKSTDYRDTYDASLLQGVPRSLNRDPLGL
                     KADNLPFQGTDIWTLYELSWLNAKGLPQVAVGHVELDYTSVNLIESKSFKLYLNSFNQ
                     TRFNNWDEVRQTLERDLSTCAQGEVSVALYRLDELEGQPIGHFNGTCIDDQDITIDNY
                     EFTTDYLENATSGEKVVEETLVSHLLKSNCLITHQPDWGSIQIQYRGRQIDREKLLRY
                     LVSFRHHNEFHEQCVERIFNDLLRFCQPEKLSVYARYTRRGGLDINPWRSNSDFVPST
                     TRLVRQ"
     gene            4017. .5381
                     /gene="ECs3655"
     CDS             4017. .5381
                     /gene="ECs3655"
                     /note="similar to YGDH_ECOLI gi|1789159 percent identity
                     99 in 454 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37078.1"
                     /db_xref="GI:13363127"
                     /translation="MITHISPLGSMDMLSQLEVDMLKRTASSDLYQLFRNCSLAVLNS
                     GSLTDNSKELLSRFENFDINVLRRERGVKLELINPPEEAFVDGRIIRAIQANLFAVLR
                     DILFVYGQIHNTVRFPNLNLDNSVHITNLVFSILRNARALHVGEAPNMVVCWGGHSIN
                     ENEYLYARRVGNQLGLRELNICTGCGPGAMEAPMKGAAVSHAQQRYKDSRFIGMTEPS
                     IIAAEPPNPLVNELIIMPDIEKRLEAFVRIAHGIIIFPGGVGTAEELLYLLGILMNPA
                     NKDQVLPLILTGPKESADYFRVLDEFVVHTLGENARRHYRIIIDDAAEVARQMKKSMP
                     LVKENRRDTGDAYSFNWSMRIAPDLQMPFEPSHENMANLKLYPDQPVEVLAADLRRAF
                     SGIVAGNVKEVGIRAIEEFGPYKINGDKEIMRRMDDLLQGFVAQHRMKLPGSAYIPCY
                     EICT"
     gene            5938. .7227
                     /gene="ECs3656"
     CDS             5938. .7227
                     /gene="ECs3656"
                     /note="similar to SDAC_ECOLI gi|1789160 percent identity
                     100 in 429 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="probable serine transporter"
                     /protein_id="BAB37079.1"
                     /db_xref="GI:13363128"
                     /translation="METTQTSTIASKDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPIN
                     AGVGGMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLIT
                     LLYFFAIYPILLVYSVAITNTVESFMSHQLGMTPPPRAILSLILIVGMMTIVRFGEQM
                     IVKAMSILVFPFVGVLMLLALYLIPQWNGAALETLSLDTASATGNGLWMTLWLAIPVM
                     VFSFNHSPIISSFAVAKEEEYGDMAEQKCSKILAFAHIMMVLTVMFFVFSCVLSLTPA
                     DLAAAKEQNISILSYLANHFNAPVIAWMAPIIAIIAITKSFLGHYLGAREGFNGMVIK
```

```
                SLRGKGKSIEINKLNRITALFMLVTTWIVATLNPSILGMIETLGGPIIAMILFLMPMY
                AIQKVPAMRKYSGHISNVFVVVMGLIAISAIFYSLFS"
     gene       7285. .8652
                /gene="ECs3657"
     CDS        7285. .8652
                /gene="ECs3657"
                /note="similar to SDAA_ECOLI gi|1788116 percent identity
                77 in 455 aa, similar to SDAB_ECOLI gi|1789161 percent
                identity 99 in 455 aa (Conserved in E.coli K-12)"

  Query Match           98.9%;  Score 1142.2;  DB 1;  Length 266658;
  Best Local Similarity  99.3%;  Pred. No. 1.7e-263;
  Matches 1147;  Conservative    0;  Mismatches    8;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183501 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 183560

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
             |||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183561 GACCAAATCTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 183620

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183621 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 183680

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183681 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 183740

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
             |||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db  183741 TCCGACATGGGCTTTGACGCTAACTCCTGTGCAGTTCTGAGCGCTATCGGCAAACAGTCT 183800

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183801 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 183860

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183861 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 183920

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  183921 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 183980

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
             |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db  183981 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTCCAGTATGACGACGGCAAAATCGTTGGT 184040

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184041 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 184100

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
             |||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db  184101 GAAGCGGTAATGGAAGAGATCATCAAGCCGATTCTGCCCGCTGAATGGCTGACTTCTGCC 184160

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
             |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db  184161 ACCAAATTCTTCATCAACCCGACCGGTCGTTTTGTTATCGGTGGCCCGATGGGTGACTGC 184220

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184221 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 184280

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184281 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 184340

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184341 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 184400

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184401 TACGCAATCGGCGTGGCTGAACCGACTTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 184460

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
             |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db  184461 GTGCCTTCTGAACAACTGACTCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 184520

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184521 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 184580

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  184581 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 184640

Qy    1141 GCCGGTCTGAAGTAA 1155
             |||||||||||||||
Db  184641 GCCGGTCTGAAGTAA 184655
```

RESULT 12
ECOMETX
LOCUS ECOMETX 1152 bp DNA linear BCT 31-JUL-1996
DEFINITION Escherichia coli S-adenosylmethionine synthetase II dimer subunit (metX) gene, complete cds.
ACCESSION M98266
VERSION M98266.1 GI:146850
KEYWORDS S-adenosylmethionine synthetase II; metX gene.
SOURCE Escherichia coli (strain K-12) (library: Lambda) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (bases 1 to 1152)
AUTHORS Satischandran,C., Taylor,J.C. and Markham,G.D.
TITLE Isozymes of S-adenosylmethionine synthetase are encoded by tandemly duplicated genes in Escherichia coli
JOURNAL Mol. Microbiol. 9 (4), 835-846 (1993)
MEDLINE 94049123
PUBMED 8231813

```
FEATURES             Location/Qualifiers
     source          1. .1152
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /db_xref="taxon:562"
                     /tissue_lib="Lambda"
     gene            1. .1152
                     /gene="metX"
     CDS             1. .1152
                     /gene="metX"
                     /EC_number="2.5.1.6"
                     /codon_start=1
                     /transl_table=11
                     /product="S-adenosylmethionine synthetase II"
                     /protein_id="AAB05197.1"
                     /db_xref="GI:146851"
                     /translation="MAKHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACET
                     YVKTGMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAIGKQSPD
                     INQGVDRADPLEQGAGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAEVRKNGTLRV
                     RPDAKSQVTFQYDDGKIVGIDAVVLSTQHSEEIDQKSLQEAVMEEIIKPILPAEWLTS
                     ATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTYGGMARHGGGAFSGKDPSKVDRSAAY
                     AARYVAKNIVAAGLADRCEIQVSYAIGVAEPTSIMVETFGTEKVPSEQLTLLVREFFD
                     LRPYGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAHVLPDAAGLK"
     misc_difference 50. .61
                     /gene="metX"
                     /note="different from metK"
     misc_difference 123. .135
                     /gene="metX"
                     /note="different form metK"
BASE COUNT      277 a    310 c    302 g    263 t
ORIGIN      63 min.
```

```
Query Match          97.8%;  Score 1129.4;  DB 1;  Length 1152;
Best Local Similarity  99.2%;  Pred. No. 1.6e-260;
Matches 1146;  Conservative    0;  Mismatches    6;  Indels    3;  Gaps    1;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGC---GT 477

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     478 GTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 537

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     538 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 597

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     598 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 657

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     658 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 717

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     718 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 777

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     778 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 837

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     838 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 897

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     898 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 957

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     958 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1017

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1018 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1077

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||  ||||||  ||||||
Db    1078 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCACGTGCTGCCGGATGCT 1137

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    1138 GCCGGTCTGAAGTAA 1152
```

```
RESULT 13
ECOMETK
LOCUS       ECOMETK                 1462 bp    DNA     linear   BCT 26-APR-1993
DEFINITION  E. coli metK gene coding for S-adenosylmethionine synthetase.
ACCESSION   K02129
VERSION     K02129.1  GI:146838
KEYWORDS    S-adenosylmethionine synthetase; metK gene; synthetase.
SOURCE      E. coli (K12) DNA, clone pK8.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1462)
  AUTHORS   Markham,G.D., DeParasis,J. and Gatmaitan,J.
  TITLE     The sequence of metK, the structural gene for S-adenosylmethionine
            synthetase in Escherichia coli
  JOURNAL   J. Biol. Chem. 259 (23), 14505-14507 (1984)
  MEDLINE   85054924
   PUBMED   6094561
COMMENT     Sequence furnished by Markham (07-APR-1985) on floppy disk. [1]
            notes a potential ribosome-binding site at 72-77.
FEATURES             Location/Qualifiers
     source          1. .1462
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     gene            86. .1240
                     /gene="metK"
     CDS             86. .1240
                     /gene="metK"
                     /codon_start=1
                     /transl_table=11
                     /product="S-adenosylmethionine synthetase"
                     /protein_id="AAA24164.1"
                     /db_xref="GI:146839"
                     /translation="MAKHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACET
                     YVKTGIGFSWRRNHHQRPWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAIGKQSP
                     DINQGVDRADPLEQGAGDQGLDVSATQLMKPTCLMPAPITYAHRLVQRQAEVRKNGTL
                     RVRPDAKSQVTFSYDDQGKIVGIDAVVLSTQHSEEIDQKSLQEAVMEEIIKPILPAEWL
                     TSATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTTGGMARHGGGAFSGKDPSKVDRSA
                     AYAARYVAKNIVAAGLADRCEIQVSYAIGLAEPTSIMVETFGTEKVPSEQLTLLVREF
                     FDLRPIGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAQLLRDAAGLK"
BASE COUNT      368 a    389 c    349 g    356 t
ORIGIN      5 bp upstream of PstI site.
```

```
  Query Match           90.3%;  Score 1043.4;  DB 1;  Length 1462;
  Best Local Similarity   97.2%;  Pred. No. 7.4e-240;
  Matches 1128;  Conservative    0;  Mismatches   21;  Indels   12;  Gaps    6;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           |||||||||||||||||||||||||||||||||||| || ||||||||||||||||||||
Db      86 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAGGGCCATCCTGACAAAATTGCT 145

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     146 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 205

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCA-TGGTTTTAGTTGGCGGCGAAATCACCACCAG 179
           |||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||
Db     206 GCTTGCGAAACCTACGTAAAAACCGGCATTGGTTTTAGTTGGCGGCGAAATCACCACCAG 265

Qy     180 CG--CCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTG 237
           ||  | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    266 CGACCTTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTG 325

Qy    238 CATTCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAG 297
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    326 CATTCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAG 385

Qy    298 TCTCCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGAC 357
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    386 TCTCCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGAC 445

Qy    358 CAGGGTCT--GATGTTTGGCTACGCAACTAATGAAACCGACGTG-CTGATGCCAGCACCT 414
          ||||||||  || ||||||||||||||||||||||||||||| ||||||||||||||||
Db    446 CAGGGTCTTGATGTTTCGGCTACGCAACTAATGAAACCGACGTGCCTGATGCCAGCACCT 505

Qy    415 ATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTG 474
          |||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    506 ATCACCTATGCCCACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTG 565

Qy    475 CCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATC 534
          |   | ||||||||||||||||||||||||||||||   ||||||||||||||||||
Db    566 C---GTGTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTAGCTATGACGACGGCAAAATC 622

Qy    535 GTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCG 594
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    623 GTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCG 682

Qy    595 CTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACT 654
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    683 CTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACT 742

Qy    655 TCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    743 TCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 802

Qy    715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774
          ||||||||||| ||||||||||||||||||||   ||||||||||||||||||||||||
Db    803 GACTGCGGTCTTACTGGTCGTAAAATTATCGTTGATACTACCGGCGGCATGGCGCGTCAC 862

Qy    775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    863 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 922

Qy    835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    923 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 982

Qy    895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 954
          ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db    983 GTTTCCTACGCAATCGGCCTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 1042

Qy    955 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCA 1014
          ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db   1043 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTG---CCA 1099

Qy   1015 TACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1074
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1100 ATCGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1159

Qy   1075 GGTCACTTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGC 1134
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1160 GGTCACTTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGC 1219

Qy   1135 GATGCTGCCGGTCTGAAGTAA 1155
          |||||||||||||||||||||
Db   1220 GATGCTGCCGGTCTGAAGTAA 1240


RESULT 14
AE008842
LOCUS       AE008842               22204 bp    DNA     linear   BCT 31-JUL-2002
DEFINITION  Salmonella typhimurium LT2, section 146 of 220 of the complete
            genome.
ACCESSION   AE008842 AE006468
VERSION     AE008842.1  GI:16421636
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2.
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 22204)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 22204)
  AUTHORS   The Salmonella typhimurium Genome Sequencing Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT      Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .22204
                     /organism="Salmonella typhimurium LT2"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            complement(657. .1463)
                     /gene="STM3084"
     CDS             complement(657. .1457)
                     /gene="STM3084"
                     /note="similar to E. coli regulator for uxu operon
                     (AAC77280.1); Blastp hit to AAC77280.1 (257 aa), 39%
                     identity in aa 8 - 247"
                     /codon_start=1
                     /transl_table=11
                     /product="putative gntR family regulatory protein"
                     /protein_id="AAL21959.1"
                     /db_xref="GI:16421637"
                     /translation="MFTLPSSGAIAVYIGMCMEQILTKRRYFDIGLQIEELLYSGVFK
                     AGERLPAERELSERFQTSRTTIREAIIMLELKGVVEVKQGAGTYFIDSLEKINQKALL
                     PYSDIGPFELLQARQVIESNITGFAATQIRFNELKQLKRIIEQQEKQIGGDSDKFEEL
                     DRQFHNIIAESTQNRVLMKQSAELWRAVRTENPRWKQLNYKYLHKKELRMKWVEDHRS
                     IFLALQKRDAEQARQASWTHLENSKNELVKIFQQDDSLEDFDDFFFAT"
     RBS             complement(1458. .1463)
                     /gene="STM3084"
                     /note="putative RBS for STM3084; RegulonDB:STMS1H003125"
```

```
gene            complement(1611. .2348)
                /gene="STM3085"
CDS             complement(1611. .2342)
                /gene="STM3085"
                /codon_start=1
                /transl_table=11
                /product="putative outer membrane lipoprotein"
                /protein_id="AAL21960.1"
                /db_xref="GI:16421638"
                /translation="MSKFKFNAVIVGLFLLLGGCSSGMQSNNGSSSDVGTAWGGDVHS
                SVQSVSVERADREPAEMVIINYSTQYPSGYDKVYSIRISDLEYAVRDANFNSIPITRR
                YNASMGQWQYSIPARSGMNYQLYIRNYSHDTNYEIVATVDGLDVLNGKAGSLNHHGYI
                VNAGDSLAIKGFRKDKHTEAAFQFADVADAYAAHSAQGDVRNIGVIGFAAFALQGKAT
                NTLPPCSSQAFPADNNGYAPPPCRK"
RBS             complement(2343. .2348)
                /gene="STM3085"
                /note="putative RBS for STM3085; RegulonDB:STMS1H003126"
gene            complement(2486. .4558)
                /gene="speA"
                /note="STM3086"
CDS             complement(2486. .4462)
                /gene="speA"
                /EC_number="4.1.1.19"
                /note="similar to E. coli biosynthetic arginine
                decarboxylase (AAC75975.1); Blastp hit to AAC75975.1 (658
                aa), 97% identity in aa 1 - 658"
                /codon_start=1
                /transl_table=11
                /product="arginine decarboxylase"
                /protein_id="AAL21961.1"
                /db_xref="GI:16421639"
                /translation="MSDDMSMGSPSSAGEQGVLRSMQEVAMSSQEASKMLRTYNIAWW
                GNNYYDVNELGHISVCPDPDVPEARVDLAKLVKAREAQGQRLPALFCFPQILQHRLRS
                INAAFKRARESYGYNGDYFLVYPIKVNQHRRVIESLIHSGEPLGLEAGSKAELMAVLA
                HAGMTRSVIVCNGYKDREYIRLALIGEKMGHKVYLVIEKMSEIAIVLEEAERLNVVPR
                LGVRARLASQGSGKWQSSGGEKSKFGLAATQVLQLVETLRDAGRLDSLQLLHFHLGSQ
                MANIRDIATGVRESARFYVELHKLGVNIQCFDVGGGLGVDYEGTRSQSDCSVNYGLNE
                YANNIIWAIGDACEEHGLPHPTVITESGRAVTAHHTVLVSNIIGVERNEYTDPTAPAE
                DAPRALQNLWETWQEMHKPGTRRSLREWLHDSQMDLHDIHIGYSSGAFSLQERAWAEQ
                LYLSMCHEVQKQLDPQNRAHRPIIDELQERMADKMYVNFSLFQSMPDAWGIDQLFPVL
                PLEGLDQVPERRAVLLDITCDSDGAIDHYIDGDGIATTMPMPEYDPENPPMLGFFMVG
                AYQEILGNMHNLFGDTEAVDVFVFPDGSVEVELSDEGDTVADMLQYVQLDPKTLLTHF
                RDQVKQTDLDDALQQQFLEEFEAGLYGYTYLEDE"
RBS             complement(4467. .4472)
                /gene="speA"
                /note="putative RBS for speA; RegulonDB:STMS1H003127"
gene            complement(4471. .4602)
                /gene="yqgB"
                /note="STM3087"
CDS             complement(4471. .4602)
                /gene="yqgB"
                /note="similar to E. coli orf, hypothetical protein
                (AAC75976.1); Blastp hit to AAC75976.1 (48 aa), 73%
                identity in aa 6 - 46"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL21962.1"
                /db_xref="GI:16421640"
                /translation="MKKKPVAQAEHQRYLLENPLVYGLLSRLRIAIVVNCFTLANKN"
-10_signal      complement(4531. .4539)
                /gene="yqgB"
                /note="putative -10_signal for speA;
                RegulonDB:STMLTH004574"
-35_signal      complement(4550. .4558)
                /gene="yqgB"
                /note="putative -35_signal for speA;
                RegulonDB:STMLTH004574"
gene            4723. .4900
                /gene="yqgC"
                /note="STM3088"
RBS             4723. .4728
                /gene="yqgC"
                /note="putative RBS for yqgC; RegulonDB:STMS1H003128"
CDS             4736. .4900
                /gene="yqgC"
                /note="similar to E. coli orf, hypothetical protein
                (AAC75977.1); Blastp hit to AAC75977.1 (71 aa), 78%
                identity in aa 1 - 51"
                /codon_start=1
                /transl_table=11
                /product="putative cytoplasmic protein"
                /protein_id="AAL21963.1"
                /db_xref="GI:16421641"
                /translation="MGITSAGMQSRDADCGGHARTRTMRQIQQNTTVHYLVSPPRPPR
                KTNPQAKATF"
gene            complement(4897. .5207)
                /gene="yqgD"
                /note="STM3089"
CDS             complement(4897. .5196)
                /gene="yqgD"
                /note="similar to E. coli orf, hypothetical protein
                (AAC75978.1); Blastp hit to AAC75978.1 (83 aa), 52%
                identity in aa 1 - 78"
                /codon_start=1
                /transl_table=11
                /product="putative inner membrane protein"
                /protein_id="AAL21964.1"
                /db_xref="GI:16421642"
                /translation="MTEDSTMSLKVADTGSLSVNQYGWINIWTAILGHFFAQFPAFFE
                GRRNVGQNQAVNVSDNADIMRIYALLFFCVLGIKRDARLIQKHTTFSSCRHFQPG"
RBS             complement(5202. .5207)
                /gene="yqgD"
                /note="putative RBS for yqgD; RegulonDB:STMS1H003129"
gene            5237. .6406
                /gene="metK"
                /note="STM3090"
RBS             5237. .5242
                /gene="metK"
                /note="putative RBS for metK; RegulonDB:STMS1H003130"
CDS             5252. .6406
                /gene="metK"
                /EC_number="2.5.1.6"
                /note="AdoMet synthetase; similar to E. coli methionine
                adenosyltransferase 1 (AdoMet synthetase); methyl and
                propylamine donor, corepressor of met genes (AAC75979.1);
                Blastp hit to AAC75979.1 (384 aa), 97% identity in aa 1 -
                384"
                /codon_start=1
                /transl_table=11
                /product="methionine adenosyltransferase 1"
                /protein_id="AAL21965.1"
                /db_xref="GI:16421643"
```

Query Match 80.5%; Score 929.4; DB 1; Length 22204;
Best Local Similarity 87.8%; Pred. No. 2e-212;
Matches 1014; Conservative 0; Mismatches 141; Indels 0; Gaps 0;

```
Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGCATCCTGACAAAATTGCT 60
          |||||||||||||||||||||| ||||||| || ||||||||||||||||||||| ||
Db   5252 ATGGCAAAACACCTTTTTACGTCTGAGTCCGTATCAGAAGGCATCCTGACAAAATCGCA 5311

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          |||||||| |||||||||||| || ||||| |||||  ||||||||||||||||||||||
Db   5312 GACCAAATCTCTGATGCCGTGTTGGACGCTATCCTGCAACAGGATCCGAAAGCACGCGTC 5371

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          || || |||||||||||| ||||||||||||||||||||||||||| || ||||||||||||
Db   5372 GCCTGTGAAACCTACGTTAAAACCGGCATGGTTTTAGTTGGCGGTGAGATCACCACCAGC 5431

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          |||||||| || |||||||||||  |||||||| || || ||||||||||||||||||||
Db   5432 GCCTGGGTCGATATCGAAGAGATTACCCGTAATACGGTGCGCGAAATTGGCTATGTGCAT 5491

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          |||||||||||||||||||| ||||| || ||||| |||||||| || ||||||||||||
```

```
Db    5492 TCCGACATGGGCTTTGACGCCAACTCTTGCGCGGTACTGAGCGCAATTGGCAAACAGTCT 5551

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           || || ||||||||||||||||||||| || || |||||||||| |||||||| ||||||
Db    5552 CCGGATATCAACCAGGGCGTTGACCGCGCAGACCCGCTGGAACAAGGCGCGGGCGACCAG 5611

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           || |||||||||||||||| || |||||||||||||| |||||||| || || ||||||
Db    5612 GGCCTGATGTTTGGCTACGCGACAAATGAAACCGACGTACTGATGCCTGCGCCAATCACC 5671

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           || || ||||||||||||||||||||||||||||||||||||||||| |||| |||| |||
Db    5672 TACGCGCACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAATGGCACCCTGCCATGG 5731

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           |||| ||||| || |||||||||| ||||| ||||||||||||||||||||||||||| |||
Db    5732 TTGCGTCCGGATGCAAAAAGCCAGGTCACTTTCCAGTATGACGACGGCAAAATCGTCGGT 5791

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||| || || || || || || |||||| | ||||| |||||||| ||||||||||||
Db    5792 ATCGACGCCGTGGTTCTCTCTACGCAGCACGCAGAAGATATCGACCAAAAATCGCTGCAA 5851

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           |||||||| |||||||||||||||| || |||||||  ||||||| | | | |  ||
Db    5852 GAAGCGGTGATGGAAGAGATCATCAAACCCATTCTGCCGTCTGAATGGTTAAATACGTCC 5911

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           || || || || ||||||| |||| |||| ||||||| |||| |||| || |||
Db    5912 ACTAAGTTTTTTATCAACCCAACCGGGCGTTTTGTTATCGGCGGCCCGATGGGCGATTGC 5971

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           |||||||| |||||||| || |||||||||||||||||||| |||||||||||||| |||
Db    5972 GGTCTGACCGGTCGTAAGATCATCGTTGATACCTACGGCGGTATGGCGCGTCACGGCGGC 6031

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           || || ||||| |||||||||||| || ||||| |||||||| ||||| |||| || |||
Db    6032 GGCGCCTTCTCCGGTAAAGATCCGTCTAAAGTTGACCGTTCTGCAGCTTACGCTGCGCGT 6091

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           |||||||||||||||||||| || || || ||||| ||||| |||||||| ||||| |||
Db    6092 TATGTCGCGAAAAACATCGTGGCGGCAGGTCTGGCGGATCGCTGTGAAATCCAGGTCTCC 6151

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||| |||||||| || || ||||| |||||||||| ||||| |||||||||| |||
Db    6152 TACGCTATCGGCGTAGCGGAGCCGACGTCCATCATGGTGGAAACGTTCGGTACTGAAAAA 6211

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           |||||| |||||||  |||  |||||||| || || ||||||||||||||||| |||||
Db    6212 GTGCCTGCTGAACAGTTGATTCTGCTGGTGCGCGAATTCTTCGACCTGCGCCCGTACGGC 6271

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           |||||||||||||||||||||||||||||||||||||||||||||| || || ||||| |||
Db    6272 TTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACTGCGGCTTACGGCCAC 6331

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           || ||||| ||| | ||||| |||||||||||||||||||||| ||||||||||||||||
Db    6332 TTCGGTCGCGAAAACTTCCCATGGGAAAAAACCGACAAAGCGCAACTGCTGCGCGATGCT 6391

Qy    1141 GCCGGTCTGAAGTAA 1155
           ||||||||||| |||
Db    6392 GCCGGTCTGAAATAA 6406


RESULT 15
AL627277
LOCUS       AL627277               230050 bp    DNA     linear   BCT 06-JUN-2002
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 13/20.
ACCESSION   AL627277 AL513382
VERSION     AL627277.1  GI:16504049
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi.
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 230050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 230050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonella
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .230050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            309. .1598
                     /gene="STY3109"
                     /note="sdaC"
     CDS             309. .1598
                     /gene="STY3109"
                     /note="Fasta hit to Q46940 (409 aa), 40% identity in 416
                     aa overlap
                     Fasta hit to TDCC_ECOLI (443 aa), 50% identity in 442 aa
                     overlap
                     Orthologue of E. coli sdaC (SDAC_ECOLI); Fasta hit to
                     SDAC_ECOLI (429 aa), 95% identity in 429 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative serine transporter"
                     /protein_id="CAD02796.1"
                     /db_xref="GI:16504050"
                     /db_xref="SPTREMBL:Q8Z435"
                     /translation="MKTTQTSTIASIDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPIN
                     AGVGGMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLIT
                     LLYFFAIYPILLVYSVAITNTVESFLTHQLAINPPPRAILSLILIVGMMTIVRFGEQM
                     IVKAMSILVFPFVAALMLLALYLIPQWNGAALETLSFDSAASTGNGLWMTLWLAIPVM
                     VFSFNHSPIISSFAVAKREEYGEGAEKKCSKILAFAHIMMVLTVMFFVFSCVLSLTPA
                     DLAAAKEQNISILSYLANHFNAPIIAWMAPIIAMIAITKSFLGHYLGAREGFNGMVIK
                     SLRGKGKSIEINKLNKITALFMLVTTWIVATLNPSILGMIETLGGPIIAMILFLMPMY
                     AIQKVPAMRKYSGHISNVFVVIMGLIAISAIFYSLFS"
     gene            1657. .3024
                     /gene="STY3110"
                     /note="sdaB"
     CDS             1657. .3024
                     /gene="STY3110"
                     /note="Fasta hit to TDCG_ECOLI (454 aa), 73% identity in
                     454 aa overlap
                     Fasta hit to SDHL_ECOLI (454 aa), 77% identity in 455 aa
                     overlap
                     Orthologue of E. coli sdaB (SDHM_ECOLI); Fasta hit to
                     SDHM_ECOLI (455 aa), 94% identity in 455 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="L-serine dehydratase 2 (L-serine deaminase 2)"
                     /protein_id="CAD02797.1"
```

```
                /db_xref="GI:16504051"
                /db_xref="SPTREMBL:Q8Z434"
                /translation="MISVFDIFKIGIGPSSSHTVGPMKAGKQFTDDLIARHILTDVTR
                VVVDVYGSLSLTGKGHHTDIAIIMGLAGNLPDTVDIDAIPSFIQDVNTHGRLLLANGQ
                HEVEFPVDKCMNFHADNLSLHENGMRITALAGNKVLYSQTYYSIGGGFIVDEEHFGLT
                NSEPVNVPYPYKTAADLQRHCQETGLSLSGLMMQNELALHSKEALEQHFAKVWEVMRS
                GIERGITTEGVLPGKLRVPRRAAALRRMLVSQDKTTTDPMAVVDWINMFALAVNEENA
                AGGRVVTAPTNGACGIVPAVLAYYDKFIREVNANSLARYMLVASAIGSLYKMNASISG
                AEVGCQGEVGVACSMAAAGLAELLGGSPAQVCIAAEIAMEHNLGLTCDPVAGQVQVPC
                IERNAIASVKAVNAARMALRRTSEPRVCLDKVIETMYETGKDMNAKYRETSRGGLAMK
                IVTCD"
gene            3135. .3950
                /gene="STY3111"
                /note="exo"
CDS             3135. .3950
                /gene="STY3111"
                /note="Orthologue of E. coli exo (EX9_ECOLI); Fasta hit to
                EX9_ECOLI (281 aa), 88% identity in 264 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="5'-3' exonuclease"
                /protein_id="CAD02798.1"
                /db_xref="GI:16504052"
                /db_xref="SPTREMBL:Q8Z433"
                /translation="MLASPRRECYPTRRFVRENTVAAHLLIVDALNLIRRIHAVQGSP
                CVETCQHALDQLIIHSQPTHAVAVFDDDARSSGWRHQRLPDYKAGRPPMPDDLHNEMP
                ALRAAFEQRGVRCWASDGNEADDLAATLALKVTEAGHQATIVSTDKGYCQLLSPGLRI
                RDYFQKRWLDAPFIEKEFGVLPRQLPDYWGLAGISSSKVPGVAGIGPKSATQLLVQFQ
                NLEGIYAHLDEVPEKWRKKLETHKEMAFLCRDIARLQTDLHIDGNLQQLRLVR"
misc_feature    3201. .3947
                /gene="STY3111"
                /note="Pfam match to entry PF01367 5_3_exonuclease, 5'-3'
                exonuclease, score 120.40, E-value 3.3e-32"
gene            complement(4042. .5190)
                /gene="STY3112"
                /note="fucO"
CDS             complement(4042. .5190)
                /gene="STY3112"
                /note="Fasta hit to EUTG_ECOLI (395 aa), 40% identity in
                372 aa overlap
                Fasta hit to ADH2_ECOLI (383 aa), 44% identity in 371 aa
                overlap
                Orthologue of E. coli fucO (FUCO_ECOLI); Fasta hit to
                FUCO_ECOLI (383 aa), 91% identity in 382 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="1,2-propanediol oxidoreductase (lactaldehyde
                reductase)"
                /protein_id="CAD02799.1"
                /db_xref="GI:16504053"
                /db_xref="SPTREMBL:Q8Z432"
                /translation="MANRMILNETAWFGRGAVDALTDEVTRRGYHKALIVTDKTLVQC
                GVVDKVTSRMDAAGLAWEIYAGVIPNPTISVVQEGLKVFTQSGADYLIAIGGGSPQDT
                CKAVGIISNNPEFADVRSLEGLSPTRKPSVPIMAIPTTAGTAAEVTINYVITDEEKRR
                KFVCVDPHDIPQAAFIDADMMDGMPPALKAATGVDALTHAIEGYITRAAWALTDALHI
                KAIEIIAGALRGAVAGEKEAGEAMALGQYVAGMGFSNVGLGLVHGMAHPLGAFYNTPH
                GVANAILLPHVMRFNAGSTNEKFRDIARAMGVKVEGLSLEEARNAAVEAVFTLNRDVG
                IPLHLRDVGVRKEDIPALAQAAFDDVCTGGNPREASLADIVELYHTAW"
misc_feature    complement(4069. .5169)
                /gene="STY3112"
                /note="Pfam match to entry PF00465 Fe-ADH, Iron-containing
                alcohol dehydrogenases, score 655.70, E-value 2.4e-193"
misc_feature    complement(4354. .4416)
                /gene="STY3112"
                /note="PS00060 Iron-containing alcohol dehydrogenases
                signature 2"
misc_feature    complement(4585. .4671)
                /gene="STY3112"
                /note="PS00913 Iron-containing alcohol dehydrogenases
                signature 1"
gene            complement(5207. .5854)
                /gene="STY3113"
                /note="fucA"
CDS             complement(5207. .5854)
                /gene="STY3113"
                /note="Fasta hit to YGBL_ECOLI (212 aa), 34% identity in
                179 aa overlap
                Orthologue of E. coli fucA (FUCA_ECOLI); Fasta hit to
                FUCA_ECOLI (215 aa), 92% identity in 215 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="fuculose-1-phosphate aldolase"
                /protein_id="CAD02800.1"
                /db_xref="GI:16504054"
                /db_xref="SPTREMBL:Q8Z431"
                /translation="MERNRLARQIIDTCLEMTRLGLNQGTAGNVSVRYQGGMLITPTG
                IPYEKLTESHIVFIDADGQHEQGKLPSSEWRFHMAAYQTRPDANAVVHNHAVHCTAVS
                ILNRPIPAIHYMIAAAGGNSIPCAPYATFGTRELSDHVAVALKNRKATLLQHHGLIAC
                EENLDKALWLAHEVEVLAQLYLSTLAIVDPVPVLDDEAIAIVLEKFKTYGLRIEE"
misc_feature    complement(5219. .5845)
                /gene="STY3113"
                /note="Pfam match to entry PF00596 Aldolase_II, Class II
                Aldolase and Adducin N-terminal domain., score 337.10,
                E-value 2e-97"
gene            6410. .7726
                /gene="STY3115"
                /note="fucP"
CDS             6410. .7726
                /gene="STY3115"
                /note="Orthologue of E. coli fucP (FUCP_ECOLI); Fasta hit
                to FUCP_ECOLI (438 aa), 92% identity in 438 aa overlap"
                /codon_start=1
                /transl_table=11
                /product="L-fucose permease"
                /protein_id="CAD02801.1"
                /db_xref="GI:16504055"
                /db_xref="SPTREMBL:Q8Z430"
                /translation="MGNTTIQTQSFRAVDAEQSKSKRYIIPFALLCSLFFLWAVANNL
                NDILLPQFQQAFTLTNFQAGLIQSAFYFGYFVIPIPAGILMKKLSYKAGIITGLFLYA
                FGAALFWPAAEIMNYTLFLIGLFIIAAGLGCLETAANPFVTALGPESGGHFRLNLAQT
                FNSFGAIIAVVFGQSLILSNVPHQSQEALDKMTPDQLSAYKHSLVLSVQTPYMIIVAI
                VLVVALLIMLTKFPALQSDDHSDAKQSSFLSSLSRLIRIRHWRWAVLAQFCYVGAQTA
                CWSYLIRYAIEEIPGMTPGFAANYLTGTMVCFFIGRFTGTWLISRFAPHKVLAAYALL
                AMLLCLISAFSGGHIGLLALTLCSAFMSIQYPTIFSLGIKNLGQDTKYGSSFIVMTII
                GGGIVTPVMGFVSDAAGKIPTAELVPALCFAVIFIFARFRSQAATN"
gene            7758. .9488
                /gene="STY3116"
                /note="fucI"
CDS             7758. .9488

  Query Match             80.3%;  Score 927.8;  DB 1;  Length 230050;
  Best Local Similarity   87.7%;  Pred. No. 5.3e-212;
  Matches 1013;  Conservative    0;  Mismatches  142;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          |||||||||||||||||||||||| ||||||| || ||||||||||||||||||||| ||
Db  131377 ATGGCAAAACACCTTTTTACGTCTGAGTCCGTATCAGAAGGGCATCCTGACAAAATCGCA 131436

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||| |||||||||| || ||||| |||||  ||||||||||||||||||||||||
Db  131437 GACCAAATCTCTGATGCCGTGTTGGACGCTATCCTGCAACAGGATCCGAAAGCACGCGTC 131496

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          || || |||||||||||||||||||||||||||||||||||||||| || ||||||||||
Db  131497 GCCTGTGAAACCTACGTTAAAACCGGCATGGTTTTAGTTGGCGGTGAGATCACCACCAGC 131556

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          |||||||| || |||||||||| |||||||| || || ||||||||||||||||||||||
Db  131557 GCCTGGGTCGATATCGAAGAGATTACCCGTAATACGGTGCGCGAAATTGGCTATGTGCAT 131616

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          |||||||||||||||||||| ||||| || ||||| |||||||| || ||||||||||||
Db  131617 TCCGACATGGGCTTTGACGCCAACTCTTGCGCGGTACTGAGCGCAATTGGCAAACAGTCT 131676

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          || || |||||||||||||||||||| || || |||||||||| |||||||| ||||||
```

```
Db 131677 CCGGATATCAACCAGGGCGTTGACCGCGCAGACCCGCTGGAACAAGGCGCGGGCGACCAG 131736

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          || |||||||||||||||| || |||||||||||||| |||||||| || || ||||||
Db 131737 GGCCTGATGTTTGGCTACGCGACAAATGAAACCGACGTACTGATGCCTGCGCCAATCACC 131796

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          || || ||||||||||||||||||||||||||||||||||||||| ||||| ||||| |||
Db 131797 TACGCGCACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAATGGCACCCTGCCATGG 131856

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          |||| ||||| || |||||||||| ||||| |||||||||||||||||||||||| |||
Db 131857 TTGCGTCCGGATGCAAAAAGCCAGGTCACTTTCCAGTATGACGACGGCAAAATCGTCGGT 131916

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||| || || || || || || ||||||| | ||||| |||||||| ||||||||||||
Db 131917 ATCGACGCCGTGGTTCTCTCTACGCAGCACGCAGAAGATATCGACCAAAAATCGCTGCAA 131976

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          |||||||| |||||||||||||||| || |||||||  |||||||| | | | |  ||
Db 131977 GAAGCGGTGATGGAAGAGATCATCAAACCCATTCTGCCGTCTGAATGGTTAAATACGTCC 132036

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          || || || || |||||||| ||||| ||||| |||||||| ||||| ||||| || ||
Db 132037 ACTAAGTTTTTTATCAACCCAACCGGGCGTTTTGTTATCGGCGGCCCGATGGGCGATTGT 132096

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          |||||||| |||||||| || ||||||||||||||||||| |||||||||||||| |||
Db 132097 GGTCTGACCGGTCGTAAGATCATCGTTGATACCTACGGCGGTATGGCGCGTCACGGCGGC 132156

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          || || ||||| |||||||||| || ||||| |||||||| ||||| ||||| || |||
Db 132157 GGCGCCTTCTCCGGTAAAGATCCGTCTAAAGTTGACCGTTCTGCAGCTTACGCTGCGCGT 132216

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          |||||||||||||||||| || || || ||||| ||||| |||||||| ||||| |||
Db 132217 TATGTCGCGAAAAACATCGTGGCGGCAGGTCTGGCGGATCGCTGTGAAATCCAGGTCTCC 132276

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||| |||||||| || || ||||| |||||||||| ||||| |||||||||| |||
Db 132277 TACGCTATCGGCGTAGCGGAGCCGACGTCCATCATGGTGGAAACGTTCGGTACTGAAAAA 132336

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          |||||| |||||||  |||  |||||||| || || ||||||||||||||||| |||||
Db 132337 GTGCCTGCTGAACAGTTGATTCTGCTGGTGCGCGAATTCTTCGACCTGCGCCCGTACGGC 132396

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||| || || ||||| |||
Db 132397 TTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACTGCGGCTTACGGCCAC 132456

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          || ||||| ||| | ||||| |||||||||||||||||||||| ||||||||||||||||
Db 132457 TTCGGTCGCGAAAACTTCCCATGGGAAAAAACCGACAAAGCGCAACTGCTGCGCGATGCT 132516

Qy   1141 GCCGGTCTGAAGTAA 1155
          ||||||||||| |||
Db 132517 GCCGGTCTGAAATAA 132531
```

Search completed: April 22, 2003, 12:16:09
Job time : 4145.55 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 08:16:32 ; Search time 290.095 Seconds
(without alignments)
8966.219 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155
Sequence: 1 atggcaaaacacctttttac.....atgctgccggtctgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1155	100.0	1155	20	AAZ09784	E. coli S-adenosyl
	2	1155	100.0	1155	22	AAA99433	Met K oligonucleot
	3	1155	100.0	1155	23	AAS52561	E. coli DNA for ce
	4	1155	100.0	3720	20	AAZ09790	Plasmid pHS1 metK
	5	1155	100.0	4975	20	AAZ09792	Plasmid pHS1 metK/
	6	1155	100.0	12354	22	AAS46243	DNA encoding novel
	7	1155	100.0	12354	22	AAS46261	DNA encoding novel
	8	1155	100.0	12354	22	AAS46267	DNA encoding novel
	9	926.8	80.2	1167	23	AAS56004	Salmonella typhi D
	10	815.4	70.6	2853	23	AAS85725	DNA encoding novel
	11	641.4	55.5	1155	23	AAS53462	Haemophilus influe
	12	586.2	50.8	1191	23	AAS54057	Pseudomonas aerugi
	13	533.4	46.2	2577	23	AAS90256	DNA encoding novel
	14	533.4	46.2	2577	23	AAS94403	DNA encoding novel
	15	505.8	43.8	3219	21	AAZ50961	E. halochloris bet
	16	503.4	43.6	56485	21	AAA81476	N. meningitidis pa
c	17	503.4	43.6	349980	21	AAF21612	Neisseria meningit
c	18	503.4	43.6	837096	21	AAA81489	N. meningitidis pa
	19	463.4	40.1	1149	22	ABA89246	Escherichia coli p
	20	463.4	40.1	27324	22	ABA89226	Escherichia coli p
	21	460	39.8	62909	22	AAF23545	Genomic fragment #
	22	391.4	33.9	1203	24	ABK73681	Bacillus lichenifo
	23	375.4	32.5	1191	23	AAS51934	Staphylococcus aur
	24	375.4	32.5	1296	23	AAS51546	Staphylococcus aur
	25	375.4	32.5	1296	23	AAS55121	Staphylococcus aur
	26	374	32.4	4848	24	AAD22684	Streptomyces fradi
	27	374	32.4	4848	24	AAD22686	Streptomyces fradi
c	28	371	32.1	2365589	24	ABA90521	Genomic sequence o
	29	370.6	32.1	1185	23	AAS53032	Enterococcus faeca
	30	369.8	32.0	1191	23	AAS55678	Streptococcus pneu
	31	369.8	32.0	5392	15	AAQ64201	Sequence comprisin
	32	369.2	32.0	1158	23	AAS51241	Enterococcus faeca
	33	368.2	31.9	1191	21	AAA15552	Streptococcus pneu
	34	368.2	31.9	1191	23	AAS55856	Streptococcus pneu
	35	368.2	31.9	1208	15	AAQ64204	snaC gene encoding
	36	368.2	31.9	8494	19	AAV52296	Streptococcus pneu
c	37	361.8	31.3	10320	18	AAV74454	Staphylococcus aur
c	38	356.6	30.9	1377	20	AAX13609	Enterococcus faeca
	39	355.4	30.8	1194	24	ABN68257	Streptococcus poly
c	40	351.4	30.4	31702	24	ABQ67190	Listeria innocua c
	41	349.2	30.2	1248	24	ABN92777	Staphylococcus epi
c	42	347.6	30.1	1210	22	AAH54976	S. epidermidis gen
	43	333.6	28.9	16710	20	AAX20544	Polynucleotide seq
	44	332.6	28.8	1194	24	ABN68256	Streptococcus poly
	45	332.2	28.8	1283	22	AAC91225	Human methionine a

ALIGNMENTS

RESULT 1
AAZ09784
ID AAZ09784 standard; DNA; 1155 BP.
XX
AC AAZ09784;
XX
DT 23-NOV-1999 (first entry)
XX
DE E. coli S-adenosyl methionine synthase SAM DNA.
XX
KW SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW bioS3; cofactor; decarboxylation; Vitamin H; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..1155
FT /*tag= a
FT /product= "S-adenosyl methionine synthase"
XX
PN DE19806872-A1.
XX
PD 26-AUG-1999.
XX
PF 19-FEB-1998; 98DE-1006872.
XX
PR 19-FEB-1998; 98DE-1006872.
XX
PA (BADI) BASF AG.
XX
PI Schroeder H;
XX

```
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33263.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
XX
PS   Claim 1; Page 8-10; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the Escherichia coli SAM protein which is used in the method of the
CC   invention.
XX
SQ   Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;

  Query Match             100.0%;  Score 1155;  DB 20;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Qy    1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db    1141 GCCGGTCTGAAGTAA 1155
```

```
RESULT 2
AAA99433
ID   AAA99433 standard; DNA; 1155 BP.
XX
AC   AAA99433;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Met K oligonucleotide SEQ ID 17 used in preparation of L-methionine.
XX
KW   L-methionine production; Met; Escherichia coli; ds.
XX
OS   Escherichia coli.
XX
PN   JP2000139471-A.
XX
PD   23-MAY-2000.
XX
PF   17-NOV-1998;   98JP-0326717.
XX
PR   17-NOV-1998;   98JP-0326717.
XX
PA   (AJIN ) AJINOMOTO KK.
XX
DR   WPI; 2001-018703/03.
DR   P-PSDB; AAB26814.
XX
PT   Fermentative process for preparation of L-methionine (Met), comprises
PT   using a modified Met producing microorganism, particularly an
PT   Escherichia species   -
XX
PS   Example 2; Page 17-18; 23pp; Japanese.
XX
CC   The invention relates to a process for preparing L-methionine (Met)
CC   using a modified Met producing organism. The modified microorganism has
CC   its Met biosynthetic system repressor deleted, and has enhanced
CC   homoserine transsuccunylase activity. The organism optionally has an
CC   attenuated intracellular S-adenosyl methionine synthetase (SAM). The
CC   process is used for the preparation of L-methionine. Sequences AAA99416
```

```
CC   and AAA99433 encode proteins represented in AAB26816 and AAB26814
CC   respectively, which are involved in the Escherichia coli Met biosynthetic
CC   pathway. AAA99417-A99441 (excluding AAA99433) represent primers, all
CC   sequences are used in the production of the modified Met producing
CC   organism of the invention.
XX
SQ   Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;

  Query Match             100.0%;  Score 1155;  DB 22;  Length 1155;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Qy    1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db    1141 GCCGGTCTGAAGTAA 1155


RESULT 3
AAS52561
ID   AAS52561 standard; DNA; 1155 BP.
XX
AC   AAS52561;
XX
DT   13-FEB-2002  (first entry)
XX
DE   E. coli DNA for cellular proliferation protein #283.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Escherichia coli.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU34702.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 6198; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
```

```
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1155 BP; 277 A; 312 C; 303 G; 263 T; 0 other;

  Query Match           100.0%;  Score 1155;  DB 23;  Length 1155;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    1141 GCCGGTCTGAAGTAA 1155


RESULT 4
AAZ09790
ID   AAZ09790 standard; DNA; 3720 BP.
XX
AC   AAZ09790;
XX
DT   23-NOV-1999  (first entry)
XX
DE   Plasmid pHS1 metK DNA.
XX
KW   SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW   bioS3; cofactor; decarboxylation; Vitamin H; metK; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             530..1684
FT                   /*tag= a
FT                   /product= "metK"
XX
PN   DE19806872-A1.
XX
PD   26-AUG-1999.
XX
PF   19-FEB-1998;   98DE-1006872.
XX
PR   19-FEB-1998;   98DE-1006872.
XX
PA   (BADI ) BASF AG.
XX
PI   Schroeder H;
XX
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33267.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
XX
PS   Example 1c; Page 25-28; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the metK protein found in plasmid pHS1 which is used in the method of
```

```
CC   the invention.
XX
SQ   Sequence 3720 BP; 932 A; 1019 C; 954 G; 815 T; 0 other;

  Query Match             100.0%;  Score 1155;  DB 20;  Length 3720;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     530 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 589

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     590 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 649

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     650 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 709

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     710 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 769

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     770 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 829

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     830 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 889

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     890 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 949

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     950 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 1009

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 1069

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1070 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 1129

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1130 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 1189

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1190 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 1249

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1250 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 1309

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1310 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 1369

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1370 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 1429

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1430 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 1489

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1490 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1549

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1550 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1609

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1610 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1669

Qy    1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db    1670 GCCGGTCTGAAGTAA 1684


RESULT 5
AAZ09792
ID   AAZ09792 standard; DNA; 4975 BP.
XX
AC   AAZ09792;
XX
DT   23-NOV-1999  (first entry)
XX
DE   Plasmid pHS1 metK/bioS1 fusion protein DNA.
XX
KW   SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW   bioS3; cofactor; decarboxylation; Vitamin H; metK; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             530..1684
FT                   /*tag= a
FT                   /product= "metK"
FT   CDS             1782..2987
FT                   /*tag= b
FT                   /product= "bioS1"
XX
PN   DE19806872-A1.
XX
PD   26-AUG-1999.
XX
PF   19-FEB-1998;   98DE-1006872.
XX
PR   19-FEB-1998;   98DE-1006872.
XX
PA   (BADI ) BASF AG.
XX
PI   Schroeder H;
XX
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33269, AAY33270.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
XX
PS   Example 3; Page 36-41; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the metK/bioS1 fusion protein found in plasmid pHS1 which is used in the
CC   method of the invention.
XX
SQ   Sequence 4975 BP; 1198 A; 1350 C; 1329 G; 1098 T; 0 other;
```

```
  Query Match             100.0%;  Score 1155;  DB 20;  Length 4975;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     530 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 589

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     590 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 649

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     650 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 709

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     710 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 769

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     770 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 829

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     830 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 889

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     890 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 949

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     950 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 1009

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1010 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 1069

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1070 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 1129

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1130 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 1189

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1190 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 1249

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1250 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 1309

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1310 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 1369

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1370 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 1429

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1430 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 1489

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1490 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1549

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1550 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1609

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1610 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1669

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    1670 GCCGGTCTGAAGTAA 1684


RESULT 6
AAS46243
ID   AAS46243 standard; DNA; 12354 BP.
XX
AC   AAS46243;
XX
DT   18-DEC-2001  (first entry)
XX
DE   DNA encoding novel mar regulated protein (NIMR) #12.
XX
KW   mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
OS   Escherichia coli.
XX
PN   WO200170776-A2.
XX
PD   27-SEP-2001.
XX
PF   08-MAR-2001; 2001WO-US07478.
XX
PR   10-MAR-2000; 2000US-188362P.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Levy SB,  Barbosa TM,  Alekshun MN;
XX
DR   WPI; 2001-602769/68.
DR   P-PSDB; AAU29343.
XX
PT   Identifying compounds that modulate a newly identified mar regulated
PT   polypeptide activity, useful as antimicrobial compounds, involves
PT   contacting the polypeptide with a test compound -
XX
PS   Disclosure; Page 224-232; 526pp; English.
XX
CC   The invention relates to a method of identifying compounds that modulate
CC   a newly identified mar regulated (NIMR) polypeptide activity. The method
CC   comprises contacting an NIMR polypeptide with a test compound under
CC   interaction conditions, determining the ability of the compound to
CC   modulate the activity or expression of the polypeptide, and selecting the
CC   modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC   of microbial infections, and in screening for modulators of NIMR
CC   expression and activity. These modulators can be used to reduce the
CC   infectivity of a microbe on a surface, and the virulence of a microbe in
CC   a subject suffering from an infection. AAS46232-AAS46278 represent
CC   Escherichia coli NIMR coding sequences of the invention.
XX
SQ   Sequence 12354 BP; 3031 A; 3249 C; 3190 G; 2884 T; 0 other;

  Query Match             100.0%;  Score 1155;  DB 22;  Length 12354;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5024 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 5083

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5084 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 5143
```

```
Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5144 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 5203

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5204 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 5263

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5264 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 5323

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5324 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 5383

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 5443

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5444 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 5503

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5504 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 5563

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5564 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 5623

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5624 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 5683

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5684 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5744 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 5803

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5804 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 5863

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5864 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 5923

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5924 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 5983

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5984 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 6043

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6044 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 6103

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6104 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 6163

Qy   1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db   6164 GCCGGTCTGAAGTAA 6178
```

```
RESULT 7
AAS46261
ID   AAS46261 standard; DNA; 12354 BP.
XX
AC   AAS46261;
XX
DT   18-DEC-2001  (first entry)
XX
DE   DNA encoding novel mar regulated protein (NIMR) #30.
XX
KW   mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
OS   Escherichia coli.
XX
PN   WO200170776-A2.
XX
PD   27-SEP-2001.
XX
PF   08-MAR-2001; 2001WO-US07478.
XX
PR   10-MAR-2000; 2000US-188362P.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Levy SB,  Barbosa TM,  Alekshun MN;
XX
DR   WPI; 2001-602769/68.
DR   P-PSDB; AAU29361.
XX
PT   Identifying compounds that modulate a newly identified mar regulated
PT   polypeptide activity, useful as antimicrobial compounds, involves
PT   contacting the polypeptide with a test compound
XX
PS   Disclosure; Page 374-382; 526pp; English.
XX
CC   The invention relates to a method of identifying compounds that modulate
CC   a newly identified mar regulated (NIMR) polypeptide activity. The method
CC   comprises contacting an NIMR polypeptide with a test compound under
CC   interaction conditions, determining the ability of the compound to
CC   modulate the activity or expression of the polypeptide, and selecting the
CC   modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC   of microbial infections, and in screening for modulators of NIMR
CC   expression and activity. These modulators can be used to reduce the
CC   infectivity of a microbe on a surface, and the virulence of a microbe in
CC   a subject suffering from an infection. AAS46232-AAS46278 represent
CC   Escherichia coli NIMR coding sequences of the invention.
XX
SQ   Sequence 12354 BP; 3031 A; 3249 C; 3190 G; 2884 T; 0 other;

  Query Match           100.0%;  Score 1155;  DB 22;  Length 12354;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5024 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 5083

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5084 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATGCGAAAGCACGCGTT 5143

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5144 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 5203

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5204 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 5263

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5264 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 5323
```

```
Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5324 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 5383

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 5443

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5444 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 5503

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5504 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 5563

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5564 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 5623

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5624 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 5683

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5684 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5744 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 5803

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5804 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 5863

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5864 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 5923

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5924 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 5983

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5984 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 6043

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6044 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 6103

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6104 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 6163

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    6164 GCCGGTCTGAAGTAA 6178


RESULT 8
AAS46267
ID   AAS46267 standard; DNA; 12354 BP.
XX
AC   AAS46267;
XX
DT   18-DEC-2001  (first entry)
XX
DE   DNA encoding novel mar regulated protein (NIMR) #36.
XX
KW   mar regulated polypeptide; NIMR; microbial infection; antibacterial; ds.
XX
OS   Escherichia coli.
XX
PN   WO200170776-A2.
XX
PD   27-SEP-2001.
XX
PF   08-MAR-2001; 2001WO-US07478.
XX
PR   10-MAR-2000; 2000US-188362P.
XX
PA   (TUFT ) TUFTS COLLEGE.
XX
PI   Levy SB, Barbosa TM, Alekshun MN;
XX
DR   WPI; 2001-602769/68.
DR   P-PSDB; AAU29368.
XX
PT   Identifying compounds that modulate a newly identified mar regulated
PT   polypeptide activity, useful as antimicrobial compounds, involves
PT   contacting the polypeptide with a test compound -
XX
PS   Disclosure; Page 424-432; 526pp; English.
XX
CC   The invention relates to a method of identifying compounds that modulate
CC   a newly identified mar regulated (NIMR) polypeptide activity. The method
CC   comprises contacting an NIMR polypeptide with a test compound under
CC   interaction conditions, determining the ability of the compound to
CC   modulate the activity or expression of the polypeptide, and selecting the
CC   modulators. NIMR nucleic acids and polypeptides are used in the treatment
CC   of microbial infections, and in screening for modulators of NIMR
CC   expression and activity. These modulators can be used to reduce the
CC   infectivity of a microbe on a surface, and the virulence of a microbe in
CC   a subject suffering from an infection. AAS46232-AAS46278 represent
CC   Escherichia coli NIMR coding sequences of the invention.
XX
SQ   Sequence 12354 BP; 3031 A; 3249 C; 3190 G; 2884 T; 0 other;

  Query Match             100.0%;  Score 1155;  DB 22;  Length 12354;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5024 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 5083

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5084 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 5143

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGGGAAATTACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5144 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 5203

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5204 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 5263

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5264 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 5323

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5324 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 5383

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5384 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 5443

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5444 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 5503
```

```
Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5504 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 5563

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5564 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 5623

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5624 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 5683

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5684 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 5743

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5744 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 5803

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5804 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 5863

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5864 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 5923

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5924 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 5983

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5984 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 6043

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6044 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 6103

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6104 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 6163

Qy   1141 GCCGGTCTGAAGTAA 1155
          |||||||||||||||
Db   6164 GCCGGTCTGAAGTAA 6178
```

RESULT 9
AAS56004
ID AAS56004 standard; DNA; 1167 BP.
XX
AC AAS56004;
XX
DT 13-FEB-2002 (first entry)
XX
DE Salmonella typhi DNA for cellular proliferation protein #37.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU38145.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 9641; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1167 BP; 284 A; 309 C; 318 G; 255 T; 1 other;

Query Match 80.2%; Score 926.8; DB 23; Length 1167;
Best Local Similarity 87.6%; Pred. No. 1e-272;
Matches 1012; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

```
Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          |||||||||||||||||||||| |||||||| || ||||||||||||||||||| ||
Db     13 ATGGCAAAACACCTTTTTACGTCTGAGTCCGTATCAGAAGGGCATCCTGACAAAATCGCA 72

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          |||||||| |||||||||||| || ||||| |||||  |||||||||||||||||||||||
Db     73 GACCAAATCTCTGATGCCGTGTTGGACGCTATCCTGCAACAGGATCCGAAAGCACGCGTC 132

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          || || |||||||||||| |||||||||||||||||||||||||| || ||||||||||||
Db    133 GCCTGTGAAACCTACGTTAAAACCGGCATGGTTTTAGTTGGCGGTGAGATCACCACCAGC 192

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          |||||||| || |||||||||||| |||||||| || || ||||||||||||||||||||||
Db    193 GCCTGGGTCGATATCGAAGAGATTACCCGTAATACGGTGCGCGAAATTGGCTATGTGCAT 252

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          |||||||||||||||||||| ||||| || ||||| |||||||| || ||||||||||||
Db    253 TCCGACATGGGCTTTGACGCCAACTCTTGCGCGGTACTGAGCGCAATTGGCAAACAGTCT 312

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          || || ||||||||||||||||||| || || |||||||||||| |||||||| ||||||
Db    313 CCGGATATCAACCAGGGCGTTGACCGCGCAGACCCGCTGGAACAAGGCGCGGGCGACCAG 372

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          || |||||||||||||||| || |||||||||||||| |||||||| || || ||||||
Db    373 GGCCTGATGTTTGGCTACGCGACAAATGAAACCGACGTACTGATGCCTGCGCCAATCACC 432

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
```

```
Db    433 TACGCGCACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAATGGCACCCTGCCATGG 492

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540

Db    493 TTGCGTCCGGATGCAAAAAGCCAGGTCACTTTCCAGTATGACGACGGCAAAATCGTCGGT 552

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Db    553 ATCGACGCCGTGGTTCTCTCTACGCAGCACGCAGAAGATATCGACCAAAAATCGCTGCAA 612

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Db    613 GAAGCGGTGATGGAAGAGATCATCAAACCCATTCTGCCGTCTGAATGGTTAAATACGTCC 672

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Db    673 ACTAAGTTTTTTATCAACCCAACCGGGCGTTTTGTTATCGGCGGCCCGATGGGCGATTGT 732

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Db    733 GGTCTGACCGGTCGTAAGATCATCGTTGATACCTACGGCGGTATGGCGCGTCACGGCGGC 792

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Db    793 GGCGCCTTCTCCGGTAAAGATCCGTCTAAAGTTGACCGTTCTGCAGCTTACGCTGCGCGT 852

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Db    853 TATGTCGCGAAAAACATCGTGGCGGCAGGTCTGGCGGATCGCTGTGAAATCCAGGTCTCC 912

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Db    913 TACGCTATCGGCGTAGCGGAGCCGACGTCCATCATGGTGGAAACGTTCGGTACTGAAAAA 972

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Db    973 GTGCCTGCTGAACAGTTGATTCTGCTGGTGCGCGAATTCTTCGACCTGCGCCNGTACGGC 1032

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Db   1033 TTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACTGCGGCTTACGGCCAC 1092

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Db   1093 TTCGGTCGCGAAAACTTCCCATGGGAAAAAACCGACAAAGCGCAACTGCTGCGCGATGCT 1152

Qy   1141 GCCGGTCTGAAGTAA 1155

Db   1153 GCCGGTCTGAAATAA 1167


RESULT 10
AAS85725
ID   AAS85725 standard; cDNA; 2853 BP.
XX
AC   AAS85725;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #21529.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG21538.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 21529; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2853 BP; 693 A; 680 C; 791 G; 689 T; 0 other;

  Query Match             70.6%;  Score 815.4;  DB 23;  Length 2853;
  Best Local Similarity   89.1%;  Pred. No. 1.8e-238;
  Matches  940;  Conservative    0;  Mismatches    1;  Indels  114;  Gaps    1;

Qy     101 AGGATCCGAAAGCACGCGTTGCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTG 160

Db    1913 AAGATCCGAAAGCACGCGTTGCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTG 1972

Qy     161 GCGGCGAAATCACCACCAGCGCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTC 220

Db    1973 GCGGCGAAATCACCACCAGCGCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTC 2032

Qy     221 GCGAAATTGGCTATGTGCATTCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGA 280

Db    2033 GCGAAATTGGCTATGTGCATTCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGA 2092

Qy     281 GCGCTATCGGCAAACAGTCTCCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGG 340

Db    2093 GCGCTATCGGCAAACAGTCTCCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGG 2152

Qy     341 AACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGC 400

Db    2153 AACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGC 2212

Qy     401 TGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTA 460

Db    2213 TGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTA 2272

Qy     461 AAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATG 520

Db    2273 AAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATG 2332
```

```
Qy    521 ACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGA 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2333 ACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGA 2392

Qy    581 TCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCG 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2393 TCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCG 2452

Qy    641 CTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCG 700
          |||||||||||||||||||||||||||||||||||||||||||
Db   2453 CTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACC---------------- 2496

Qy    701 GTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCG 760

Db   2497 ------------------------------------------------------------ 2496

Qy    761 GCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTT 820
                                                  ||||||||||||||||||||
Db   2497 ------------------------------------GATCCATCAAAAGTGGACCGTT 2518

Qy    821 CCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATC 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2519 CCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATC 2578

Qy    881 GTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAG 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2579 GTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAG 2638

Qy    941 AAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCT 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2639 AAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCT 2698

Qy   1001 TCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAG 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2699 TCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAG 2758

Qy   1061 AAACCGCAGCATACGGTCACTTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAG 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2759 AAACCGCAGCATACGGTCACTTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAG 2818

Qy   1121 CGCAGCTGCTGCGCGATGCTGCCGGTCTGAAGTAA 1155
          |||||||||||||||||||||||||||||||||||
Db   2819 CGCAGCTGCTGCGCGATGCTGCCGGTCTGAAGTAA 2853
```

RESULT 11
AAS53462
ID AAS53462 standard; DNA; 1155 BP.
XX
AC AAS53462;
XX
DT 13-FEB-2002 (first entry)
XX
DE Haemophilus influenzae DNA for cellular proliferation protein #244.
XX
KW Antisense; ds; prokaryotic cellular proliferation gene;
KW antibiotic; antibacterial; drug design.
XX
OS Haemophilus influenzae.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR P-PSDB; AAU35603.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Claim 27; Seq ID No 7099; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence encodes an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1155 BP; 348 A; 190 C; 267 G; 350 T; 0 other;

Query Match 55.5%; Score 641.4; DB 23; Length 1155;
Best Local Similarity 72.2%; Pred. No. 1.8e-185;
Matches 834; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

```
Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||| | |   |  | ||||| || || || ||||||||||||||||| || ||||||||
Db      1 ATGTCTAGTTATTTATTTACATCTGAATCAGTCTCTGAAGGGCATCCAGATAAAATTGCC 60

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          || ||||||||||||||| ||  | || |  |||||  ||||| ||||| ||||||||| ||
Db     61 GATCAAATTTCTGATGCGGTACTTGATGAAATCCTAAAACAAGATCCAAAAGCACGAGTA 120

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||| ||||| |||||||||||| || ||||  || |||||| || |||| || ||
Db    121 GCTTGTGAAACTTACGTAAAAACTGGTATGGCATTGGTTGGTGGTGAAATTACAACATCA 180

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCCTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          || ||||| || || ||  |  | ||||||||     || | || ||||| |||  |||
Db    181 GCATGGGTTGATATTGAGAATTTAACCCCTAAAGTGATTTGTGATATTGGTTATGAACAT 240

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTCTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          || || |||||||||||| | | | || ||||| || || | || || || ||||| |||
Db    241 TCTGAAATGGGCTTTGATGGTCATTCTTGTGCAGTGCTTAATGCGATTGGTAAACAATCT 300

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
            | || || || || |||||||| |||| |||| | || || ||||| || |||||
Db    301 GCAGATATTAATCAAGGCGTTGATCGTGAAAATCCATTAGATCAAGGCGCAGGCGACCAA 360

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||| | |||||||| || || || |||||||| || ||  | ||||| ||| | || ||
Db    361 GGTATTATGTTTGGTTATGCCACCAATGAAACGGATGTATTAATGCCTGCAGCCATTACT 420

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          |||||||| ||| |  |  |    ||  |||||||||||||| ||| |   |  | |||
Db    421 TATGCACATCGTTTAATGGAAAAACAATCTGAAGTGCGTAAAAGCGGTAAATTAGCATGG 480

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
```

```
Db    481 TTACGCCCAGATGCGAAAAGCCAAGTTACTTTAAAATATGAAGATAATAAAATTGTTGGT 540

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Db    541 GTGGATGCGGTTGTGCTTTCTACTCAACATTCTGAAGAAGTTAGCCAAAAAGATTTACAT 600

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Db    601 GAAGGTGTGATGGAAGAAATTATCAAACCTGTGTTACCAAGTGAATGGCTTTCTAAAGAA 660

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Db    661 ACAAAATTCTTCATTAACCCGACTGGTCGTTTTGTTATCGGTGGGCCAATGGGCGATTGT 720

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Db    721 GGTTTAACTGGCCGTAAAATCATCGTGGATACTTACGGTGGTGCCGCTCGTCACGGTGGT 780

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Db    781 GGCGCATTCTCAGGAAAAGAACCATCTAAAGTTGACCGTTCTGCGGCTTATGCTGCACGT 840

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Db    841 TATGTCGCAAAAAATATTGTGGCAGCAGGTTTAGCAGATCGTTGTGAAATCCAACTTTCT 900

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Db    901 TATGCGATTGGTGTGGCTGAACCAACATCTATCATGGTCGAAACCTTTGGAACTGGAAAA 960

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Db    961 GTTGCGAATGAATTATTAGTGTCATTAGTTCGTGAGTTTTTTGATTTACGTCCTTATGGC 1020

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Db   1021 TTAATTAAAATGTTAGATTTAATCCAACCAATCTATCGTGAAACGGCAGCTTATGGTCAT 1080

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Db   1081 TTCGGTCGTGAACAATTCCCTTGGGAAAAAGTGGATCGCGCAGAAGATTTACGTATTGCG 1140

Qy   1141 GCCGGTCTGAAGTAA 1155

Db   1141 GCAGGTTTAAAATAA 1155


RESULT 12
AAS54057
ID   AAS54057 standard; DNA; 1191 BP.
XX
AC   AAS54057;
XX
DT   13-FEB-2002  (first entry)
XX
DE   Pseudomonas aeruginosa DNA for cellular proliferation protein #188.
XX
KW   Antisense; ds; prokaryotic cellular proliferation gene;
KW   antibiotic; antibacterial; drug design.
XX
OS   Pseudomonas aeruginosa.
XX
PN   WO200170955-A2.
XX
PD   27-SEP-2001.
XX
PF   21-MAR-2001; 2001WO-US09180.
XX
PR   21-MAR-2000; 2000US-191078P.
PR   23-MAY-2000; 2000US-206848P.
PR   26-MAY-2000; 2000US-207727P.
PR   23-OCT-2000; 2000US-242578P.
PR   27-NOV-2000; 2000US-253625P.
PR   22-DEC-2000; 2000US-257931P.
PR   16-FEB-2001; 2001US-269308P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Haselbeck R,  Ohlsen KL,  Zyskind JW,  Wall D,  Trawick JD,  Carr GJ;
PI   Yamamoto RT,  Xu HH;
XX
DR   WPI; 2001-611495/70.
DR   P-PSDB; AAU36198.
XX
PT   New polynucleotides for the identification and development of
PT   antibiotics, comprise sequences of antisense nucleic acids -
XX
PS   Claim 27; Seq ID No 7694; 511pp; English.
XX
CC   The invention relates to antisense inhibitors of genes essential to
CC   prokaryotic cellular proliferation, their use in identifying the
CC   genes, their use in the discovery of novel antibiotics, the essential
CC   genes themselves and the encoded proteins. The prokaryotes used are
CC   Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC   pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC   invention is also useful for the identification of potential new targets
CC   for antibiotic development. The antisense nucleic acids can also be used
CC   to identify proteins used in proliferation, to express these proteins,
CC   and to obtain antibodies capable of binding to the expressed proteins.
CC   The proteins can be used to screen compounds in rational drug discovery
CC   programmes. The antisense nucleic acid sequence is also useful to screen
CC   for homologous nucleic acids which are required for cell proliferation in
CC   a wide variety of organisms. The present sequence encodes an
CC   essential prokaryotic cellular proliferation protein.
CC   Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1191 BP; 222 A; 420 C; 365 G; 184 T; 0 other;

  Query Match           50.8%;  Score 586.2;  DB 23;  Length 1191;
  Best Local Similarity   71.2%;  Pred. No. 1.4e-168;
  Matches  774;  Conservative    0;  Mismatches  313;  Indels    0;  Gaps    0;

Qy     12 CCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTC 71

Db     15 CGTTTTCACCTCCGAGTCCGTGTCCGAAGGCCATCCGGACAAGATCGCGGACCAGATCTC 74

Qy     72 TGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAAC 131

Db     75 CGACGCGGTGCTGGACGCCATCATCGCCAAGGACAAGTACGCTCGCGTGCCATGCGAAAC 134

Qy    132 CTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGA 191

Db    135 CCTGGTCAAGACCGGCGTGGCGATCATCGCCGGTGAAGTCACCACCTCCGCCTGGGTCGA 194

Qy    192 CATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATGGG 251

Db    195 CCTGGAAGAGCTGGTGCGCAAGGTCATCATCGACATCGGCTACGACAGCTCCGACGTCGG 254

Qy    252 CTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAA 311

Db    255 CTTCGATGGTGCCACCTGCGGCGTGCTGAACATCATCGGCAAGCAGTCGGTGGACATCAA 314

Qy    312 CCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTT 371

Db    315 CCAGGGCGTCGACCGGGCCAAGCCGGAAGACCAGGGCGCCGGCGACCAGGGCCTGATGTT 374

Qy    372 TGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCG 431

Db    375 CGGCTACGCCAGCAACGAGACCGACGTGCTGATGCCGGCGCCGATCTGTTTCTCCCACCG 434

Qy    432 TCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGA 491
```

```
Db    435 CCTGGTCGAGCGCCAGGCCGAGGCGCGCAAGTCCGGCCTGCTGCCATGGCTGCGCCCGGA 494

Qy    492 CGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGT 551

Db    495 CGCCAAGTCCCAGGTGACCTGCCGCTACGAAGGCGGCAAGGTGGTCGGCATCGATGCGGT 554

Qy    552 CGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAAT 611

Db    555 GGTGCTGTCCACCCAGCACAACCCGGAAGTCTCCTATAACGACCTGCGCGACGGCGTGAT 614

Qy    612 GGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTT 671

Db    615 GGAGCTGATCATCAAGCAGGTGCTGCCGGCCGAGCTGCTGCACAAGGACACCCAGTTCCA 674

Qy    672 CATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGG 731

Db    675 CATCAACCCGACCGGCAACTTCGTGATCGGCGGCCCGGTCGGCGACTGCGGCCTGACCGG 734

Qy    732 TCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTC 791

Db    735 ACGCAAGATCATCGTCGACTCCTACGGCGGCATGGCCCGCCACGGCGGCGGCGCGTTCTC 794

Qy    792 TGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAA 851

Db    795 CGGCAAGGACCCGTCCAAGGTCGACCGCTCCGCCGCCTACGCCGGCCGCTACGTGGCGAA 854

Qy    852 AAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGG 911

Db    855 GAACATCGTCGCCGCCGGCCTCGCCGAGCGTTGCGAGATCCAGGTTTCCTACGCCATCGG 914

Qy    912 CGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGA 971

Db    915 CGTGGCCCAGCCGACCTCGATCTCGATCAACACCTTCGGCACCGGCAAGGTCAGCGACGA 974

Qy    972 ACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGAT 1031

Db    975 GAAGATCGTGCAACTGGTCCGCGAGCACTTCGACCTGCGCCCGTACGCGATCACCAAGAT 1034

Qy   1032 GCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGA 1091

Db   1035 GCTCGACCTGCTCCACCCGATGTACCAGCCGACCGCGGCCTACGGCCACTTCGGTCGTCA 1094

Qy   1092 ACATTTC 1098

Db   1095 TCCGTTC 1101


RESULT 13
AAS90256
ID   AAS90256 standard; cDNA; 2577 BP.
XX
AC   AAS90256;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #26060.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG26069.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity
XX
PS   Claim 1; SEQ ID No 26060; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2577 BP; 629 A; 676 C; 708 G; 564 T; 0 other;

  Query Match             46.2%;  Score 533.4;  DB 23;  Length 2577;
  Best Local Similarity   99.8%;  Pred. No. 2.8e-152;
  Matches  534;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    543 CGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGA 602

Db    549 CGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGA 608

Qy    603 AGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCAC 662

Db    609 AGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCAC 668

Qy    663 CAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGG 722

Db    669 CAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGG 728

Qy    723 TCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGG 782

Db    729 TCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGG 788

Qy    783 TGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTA 842

Db    789 TGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTA 848

Qy    843 TGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTA 902

Db    849 TGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTA 908

Qy    903 CGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGT 962

Db    909 CGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGT 968

Qy    963 GCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCT 1022

Db    969 GCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCT 1028

Qy   1023 GATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGT 1077
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    1029 GATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGCT 1083


RESULT 14
AAS94403
ID   AAS94403 standard; cDNA; 2577 BP.
XX
AC   AAS94403;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #30207.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG30216.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 30207; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 2577 BP; 629 A; 676 C; 708 G; 564 T; 0 other;

  Query Match             46.2%;  Score 533.4;  DB 23;  Length 2577;
  Best Local Similarity   99.8%;  Pred. No. 2.8e-152;
  Matches  534;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy     543 CGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGA 602
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 CGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGA 608

Qy     603 AGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCAC 662
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     609 AGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCAC 668

Qy     663 CAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGG 722
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     669 CAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGG 728

Qy     723 TCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGG 782
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     729 TCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGG 788

Qy     783 TGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTA 842
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     789 TGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTA 848

Qy     843 TGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTA 902
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     849 TGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTA 908

Qy     903 CGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGT 962
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     909 CGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGT 968

Qy     963 GCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCT 1022
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     969 GCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCT 1028

Qy    1023 GATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGT 1077
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db    1029 GATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGCT 1083


RESULT 15
AAZ50961
ID   AAZ50961 standard; DNA; 3219 BP.
XX
AC   AAZ50961;
XX
DT   05-JUN-2000  (first entry)
XX
DE   E. halochloris betaine operon.
XX
KW   Betaine operon; GSMT; glycine-sarcosine methyltransferase; SDMT;
KW   sarcosine-dimethylglycine methyltransferase; animal feed;
KW   S-adenosyl methionine synthase; SAMS; tolerance; salt; freezing; cold;
KW   drought stress; pathogenesis-related protein; pH tolerance;
KW   environmental stress; inclusion body formation; ss.
XX
OS   Ectothiorhodospira halochloris.
XX
FH   Key             Location/Qualifiers
FT   CDS             221..1027
FT                   /*tag= a
FT                   /product= "Glycine-sarcosine methyltransferase"
FT   CDS             1031..1870
FT                   /*tag= b
FT                   /product= "Sarcosine-dimethylglycine methyltransferase"
FT   CDS             2006..3007
FT                   /*tag= c
FT                   /product= "S-adenosyl methionine synthase"
XX
PN   WO200011142-A2.
XX
PD   02-MAR-2000.
XX
PF   18-AUG-1999;   99WO-EP06037.
XX
PR   20-AUG-1998;   98US-0137434.
XX
PA   (CULT-) CULTOR CORP.
XX
```

PI Reinikainen T, Nyyssoelae A, Kerovuo J;
XX
DR WPI; 2000-224686/19.
DR P-PSDB; AAY70040, AAY70041, AAY70042.
XX
PT New methyltransferases, useful for e.g. producing transgenic plants
PT with increased stress tolerance, pathogen resistance or nutritional
PT value as animal feed or for improving the viability of microorganisms
PT in the food industry
XX
PS Claim 1; Figure 8; 176pp; English.
XX
CC The patent discloses the use of novel methyltransferases catalysing one
CC of the following reactions; glycine to sarcosine (GSMT), sarcosine to
CC dimethyl glycine and dimethyl glycine to betaine (SDMT) or an enzyme
CC catalysing the synthesis of S-adenosyl methionine (SAMS). The present
CC sequence is the Ectothiorhodospira halochloris betaine operon encoding
CC the methyltransferases and SAMS, which can be used to increase
CC intracellular concentrations of these compounds. The polynucleotides can
CC be used to increase salt, freezing or cold tolerance, increase
CC resistance to drought stress, pathogens or induce pathogenesis-related
CC proteins in plants. Transgenic organisms can be used as an animal feed
CC ingredient. The polynucleotides can also be used to enhance pH tolerance
CC and improve viability of organisms when subjected to environmental
CC stress. They can decrease inclusion body formation when used in
CC conjunction with polynucleotides encoding a heterologous protein.
XX
SQ Sequence 3219 BP; 745 A; 844 C; 937 G; 693 T; 0 other;

Query Match 43.8%; Score 505.8; DB 21; Length 3219;
Best Local Similarity 71.3%; Pred. No. 8.7e-144;
Matches 682; Conservative 0; Mismatches 272; Indels 3; Gaps 1;

```
Qy      9 ACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAAT 68
          | | || ||||| || ||||| |||||||||||| || || |||||||| || ||||| ||
Db   2017 ATATCTATTTACCTCTGAGTCGGTCTCTGAAGGCCACCCGGACAAAATGGCCGACCAGAT 2076

Qy     69 TTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGA 128
          ||| |||||  |  | || | | ||||    |||||||| ||  | || || |||   ||
Db   2077 TTCGGATGCACTGCTCGATGAGTTCCTGCGCCAGGATCCAAAGTCGCGGGTGGCTGCAGA 2136

Qy    129 AACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGT 188
           ||     |  | |||||||||||| |||||| |||||||| |  | ||  |||  | |
Db   2137 GACGATGATTCAGACCGGCATGGTTGTAGTTGCCGGCGAAATTAAGAGCAATGCCAAGAT 2196

Qy    189 AGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACAT 248
            | | ||   | |    ||| |    || ||||| || |||||       || || ||
Db   2197 CAATGTTGAGCCGCTGGTACGTGAAGTTGTCCGCGATATCGGCTACACCAGCTCAGATAT 2256

Qy    249 GGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACAT 308
          |||||||||| ||  || ||||||| || || | |||  ||||| | ||||| || |||||
Db   2257 GGGCTTTGATGCCGACACCTGTGCCGTACTCAACGCCCTCGGCGAGCAGTCCCCCGACAT 2316

Qy    309 CAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGAT 368
          ||| || |||||||||||| |  ||      ||| |||||||| |||||||| || |||||
Db   2317 CAATCAAGGCGTTGACCGGGAAGAGGAAGAGGAGCAGGGCGCCGGTGACCAAGGACTGAT 2376

Qy    369 GTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACA 428
          ||| || ||||| || ||||| |||||||| || ||||| ||| | |||  |||  | ||
Db   2377 GTTCGGTTACGCCACCAATGAGACCGACGTCCTCATGCCGGCAGCGATCCACTACTCGCA 2436

Qy    429 CCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCC 488
          ||  |||||  |||| ||  | ||||| |||||    |    |  |||||||| |||||||
Db   2437 CCTGCTGGTTAAGCGCCAATCCGAAGTCCGTAACTCCAAGAAGCTGCCGTGGTTGCGCCC 2496

Qy    489 GGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGC 548
          |||||| || |||||||| || ||  |||| || | ||  || ||||| |||  ||||||
Db   2497 GGACGCTAAGAGCCAGGTCACCTTCAAGTACGAAGGCGATAAGATCGTCGGTTGCGATGC 2556

Qy    549 TGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGT 608
           || ||||| ||||| |||||    ||     | || |||||  |  | || || |  ||
Db   2557 GGTGGTGCTCTCCACGCAGCATGACGAGACCGTTGATCAGAAGACCGTCCACGAGGGCGT 2616

Qy    609 AATGGAAGAGATCATCAAGCCAATTCT---GCCCGCTGAATGGCTGACTTCTGCCACCAA 665
          |||||||||||||||||||| || ||      | | |  ||||||||    |  |||||
Db   2617 GATGGAAGAGATCATCAAGCCCATCCTCGGTGACACCGGCTGGCTGACCAACGAGACCAA 2676

Qy    666 ATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCT 725
          | |  ||| |||||||||| || |||||||  ||||| ||  ||||||||||  ||  |
Db   2677 GTACCACATTAACCCGACCGGCCGCTTCGTTACTGGTGGTCCGTTGGGTGACTGTGGCTT 2736

Qy    726 GACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGC 785
          ||| || ||||| || ||||| |||||||||||||| ||||  |||||||| ||||||||
Db   2737 GACCGGCCGTAAGATCATCGTCGATACCTACGGCGGTATGGGCCGTCACGGCGGCGGTGC 2796

Qy    786 ATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGT 845
          ||||| || |||||||| || || || || || || || || || |  |  ||||||||
Db   2797 CTTCTCCGGCAAAGATCCGTCTAAGGTCGATCGCTCGGCGGCTTATGTTGGCCGTTATGT 2856

Qy    846 CGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGC 905
          || || ||||| || ||||| || |||||||| || |||||  | ||| | ||||||||
Db   2857 AGCTAAGAACATAGTGGCTGCCGGTCTGGCCGACCGCTGTGAGGTGCAGCTCTCCTACGC 2916

Qy    906 AATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGT 962
          ||||| || || || |||||||| ||| || ||||| || ||||| | || ||
Db   2917 CATCGGTGTTGCCGAGCCGACCTCGGTCAATGTCGAAACCTTTGGTACCGGCAAGGT 2973
```

Search completed: April 22, 2003, 10:11:47
Job time : 351.095 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:19:24 ; Search time 50.3875 Seconds
(without alignments)
7029.746 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155
Sequence: 1 atggcaaaacacctttttac.....atgctgccggtctgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	374	32.4	4848	4	US-08-955-957A-1	Sequence 1, Appli
	2	374	32.4	4848	4	US-08-955-957A-4	Sequence 4, Appli
	3	374	32.4	4848	4	US-08-955-957A-6	Sequence 6, Appli
	4	369.8	32.0	1208	2	US-08-403-852D-4	Sequence 4, Appli
	5	369.8	32.0	1208	3	US-08-510-646B-4	Sequence 4, Appli
	6	369.8	32.0	1208	4	US-09-231-818-4	Sequence 4, Appli
	7	369.8	32.0	5392	2	US-08-403-852D-1	Sequence 1, Appli
	8	369.8	32.0	5392	3	US-08-510-646B-1	Sequence 1, Appli
	9	369.8	32.0	5392	4	US-09-231-818-1	Sequence 1, Appli
	10	368.2	31.9	1191	4	US-09-273-686-1	Sequence 1, Appli
	11	368.2	31.9	8494	4	US-08-961-527-163	Sequence 163, App
	12	349.2	30.2	1248	4	US-09-134-001C-2240	Sequence 2240, Ap
	13	332.2	28.8	1283	4	US-09-318-448-22	Sequence 22, Appl
	14	316	27.4	3228	4	US-09-318-448-20	Sequence 20, Appl
	15	313.2	27.1	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	16	313.2	27.1	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	17	297.6	25.8	1693	3	US-09-320-878-23	Sequence 23, Appl
c	18	208	18.0	2774	4	US-09-363-189B-5	Sequence 5, Appli
c	19	169	14.6	10061	4	US-09-221-017B-672	Sequence 672, App
c	20	159.4	13.8	725	4	US-08-998-416-1020	Sequence 1020, Ap
c	21	35.4	3.1	50937	4	US-09-428-517-1	Sequence 1, Appli
	22	34.8	3.0	25165	4	US-09-453-702B-39	Sequence 39, Appl
	23	32.8	2.8	423	1	US-08-470-179-182	Sequence 182, App
c	24	31.8	2.8	5295	4	US-09-221-017B-27	Sequence 27, Appl
	25	31.8	2.8	10061	4	US-09-221-017B-672	Sequence 672, App
c	26	31.4	2.7	534	3	US-09-081-180-20	Sequence 20, Appl
c	27	31.4	2.7	534	3	US-09-040-786-20	Sequence 20, Appl
c	28	31.4	2.7	1439	4	US-09-056-556-167	Sequence 167, App
c	29	31.4	2.7	1439	4	US-09-072-596-162	Sequence 162, App
c	30	31.4	2.7	25165	4	US-09-453-702B-39	Sequence 39, Appl
c	31	31.4	2.7	30001	1	US-08-125-468-1	Sequence 1, Appli
c	32	31.4	2.7	30001	2	US-08-474-933-1	Sequence 1, Appli
	33	31	2.7	3616	4	US-09-212-247C-7	Sequence 7, Appli
	34	30.8	2.7	10216	2	US-08-875-154-1	Sequence 1, Appli
c	35	30.8	2.7	19182	2	US-08-850-880-11	Sequence 11, Appl
c	36	30.8	2.7	19182	2	US-08-944-916-11	Sequence 11, Appl
c	37	30.8	2.7	19182	4	US-09-272-432A-11	Sequence 11, Appl
	38	30.2	2.6	1434	4	US-09-491-772-2	Sequence 2, Appli
c	39	30.2	2.6	1632	1	US-08-362-232-1	Sequence 1, Appli
c	40	30.2	2.6	1632	1	US-08-814-196-1	Sequence 1, Appli
c	41	30.2	2.6	6914	1	US-08-920-812-22	Sequence 22, Appl
c	42	30.2	2.6	6914	1	US-08-920-827-22	Sequence 22, Appl
c	43	30.2	2.6	6914	1	US-08-921-177-22	Sequence 22, Appl
c	44	30.2	2.6	6914	1	US-08-362-577C-22	Sequence 22, Appl
c	45	30.2	2.6	6914	2	US-08-920-828-22	Sequence 22, Appl

ALIGNMENTS

RESULT 1
US-08-955-957A-1
; Sequence 1, Application US/08955957A
; Patent No. 6312920
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Rosteck Jr., Paul R.
; TITLE OF INVENTION: SAM Operon
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/955,957A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Webster, Thomas D.
; REGISTRATION NUMBER: 39,872
; REFERENCE/DOCKET NUMBER: P-10162
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3334
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4848 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 986..2209
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2241..3341
US-08-955-957A-1

Query Match 32.4%; Score 374; DB 4; Length 4848;
Best Local Similarity 59.7%; Pred. No. 4.2e-101;

```
  Matches  728;  Conservative    0;  Mismatches  420;  Indels   72;  Gaps    3;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||| ||   | ||| || || || ||||||||  | || || || || || ||||| || ||
Db    986 ATGTCACGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 1045

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||  ||     ||  ||||  | || ||  | ||||  | ||| |||   ||  | ||
Db   1046 GACCGGATCAGCGACACCGTCCTCGATGCCCTGCTCGCCCGGGACCCGCGGGCCAGGGTC 1105

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||   ||| |||    | |  |||||   |||      | | ||||||  |||||||| |
Db   1106 GCCGTCGAGACCCTGATCACCACCGGTCAGGTCCATATCGCCGGCGAGGTCACCACCACC 1165

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          || |  |    |||||    |  | '.  ||  |||||||  |     || ||||| |
Db   1166 GCGTACGCGCCCATCGCCCAACTGGTGCGCGACACCGTCCTGTCCATCGGCTACGACTCC 1225

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          || | || |||||| ||||    ||| || | |||   |  | | ||||||   |||||
Db   1226 TCGGCCAAGGGCTTCGACGGCGCCTCGTGCGGGGTGTCGGTCTCCATCGGCGCGCAGTCC 1285

Qy    301 CCTGACATCAACCAGGGCGTTGAC---------------------------------- 324
          || ||||||  || ||| || |||
Db   1286 CCGGACATCGCCCGGGGTGTGGACACCGCGTACGAGCGGCGGGGCGGGGGCACGGCCCCG 1345

Qy    325 ---------CGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGC 375
                  ||| ||| | |      ||||||||||| |||||||| |||||||| |||
Db   1346 GGCGGACCGGGTGACGAGCTGGACCGGCAGGGCGCGGGCGACCAGGGCCTGATGTTCGGC 1405

Qy    376 TACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTG 435
          |||||     | || |||  || ||||||||||    || |||| |   || ||||| ||
Db   1406 TACGCCTGCGACGAGACCCCCGAGCTGATGCCGCTGCCGATCAACCTGGCGCACCGGCTC 1465

Qy    436 GTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCG 495
            | ||| | | | || ||||| || ||||||||  | || |  || ||||| ||||
Db   1466 TCCCGGCGGCTGTCGGAGGTGCGGAAGAACGGCACGATCCCCTACCTCCGCCCCGACGGC 1525

Qy    496 AAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTG 555
          || | |||||| ||  |  |||| |||| || |||    ||  |  |||| | ||||||
Db   1526 AAGACCCAGGTCACCATCGAGTACGACGGCGACAAGGCGGTCCGCCTCGACACGGTCGTG 1585

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615
            | ||| | |||||| |    |  ||||||| | | ||||| |  |   | |  ||
Db   1586 GTCTCCTCCCAGCACGCCTCCGGCATCGACCTGGACTCGCTACTGGCGCCCGACATCCGC 1645

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCA-- 673
            |   ||  |        | | ||| | |  |          | || |||  |   ||
Db   1646 CGGCATGTCGTGGAGCCCGTCCTCGCCGGACTGGCCGAGGACGGCATCAAGCTGGACACC 1705

Qy    674 ------------------TCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714
                             |||| || ||||| || ||||  ||||| ||||| |||||
Db   1706 GCGGGCTACCGGCTGCTGGTCAATCCCACCGGCCGGTTCGAGATCGGCGGCCCGATGGGC 1765

Qy    715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774
          |||   || ||||| || || || || ||| | || ||||||||||||||||| || ||
Db   1766 GACGCGGGCCTGACCGGCCGCAAGATCATCATCGACACCTACGGCGGCATGGCCCGGCAT 1825

Qy    775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834
          || |||||||| ||||| || || || || || || ||||||||||| || || ||||||
Db   1826 GGCGGCGGTGCCTTCTCCGGCAAGGACCCGTCCAAGGTGGACCGTTCGGCGGCGTACGCA 1885

Qy    835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894
            || |  || || || || ||| |||| || || || ||||||   || |||||  | |||
Db   1886 ATGCGCTGGGTGGCCAAGAACGTCGTGGCGGCGGGGCTGGCCTCGCGGTGTGAGGTGCAG 1945

Qy    895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 954
          ||  ||||||| ||||||  ||| || |||  |     |  | || || || ||||| ||
Db   1946 GTCGCCTACGCCATCGGCAAGGCCGAGCCGGTCGGTCTGTTCGTGGAGACCTTCGGCACC 2005

Qy    955 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCA 1014
          |  |  |||      || |   |    | |  | |  | ||| ||||||||| |||||
Db   2006 GCCACCGTGGACGTGGAGCGGATCGAGCAGGCGATCGGCGAGGTCTTCGACCTCCGCCCG 2065

Qy   1015 TACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1074
            ||   | || | |   ||||| ||||| | ||||||||||    | ||||| || |||
Db   2066 GCCGCGATCATCCGGGATCTGGACCTGCTCCGCCCGATCTACGCCAAGACCGCCGCCTAC 2125

Qy   1075 GGTCACTTTGGTCGTGAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTG 1128
          || ||||| || || ||||      | ||| | ||||||   ||||||    |  ||| |
Db   2126 GGCCACTTCGGCCGCGAACTGCCCGAGTTCACCTGGGAACGCACCGACCGCACCGAGCAG 2185

Qy   1129 CTGCGCGATGCTGCCGGTCT 1148
          ||   ||  || ||||| ||
Db   2186 CTCATCGCCGCGGCCGGCCT 2205


RESULT 2
US-08-955-957A-4
; Sequence 4, Application US/08955957A
; Patent No. 6312920
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Rosteck Jr., Paul R.
;    TITLE OF INVENTION:  SAM Operon
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  U.S.
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0,.Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/955,957A
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Webster, Thomas D.
;      REGISTRATION NUMBER:  39,872
;      REFERENCE/DOCKET NUMBER:  P-10162
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3334
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4848 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  3338..4255
US-08-955-957A-4

  Query Match             32.4%;  Score 374;  DB 4;  Length 4848;
  Best Local Similarity   59.7%;  Pred. No. 4.2e-101;
  Matches  728;  Conservative    0;  Mismatches  420;  Indels   72;  Gaps    3;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||| ||   | ||| || || || ||||||||  | || || || || || ||||| || ||
Db    986 ATGTCACGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 1045

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||  ||     ||  ||||  | || ||  | ||||  | ||| |||   ||  | ||
Db   1046 GACCGGATCAGCGACACCGTCCTCGATGCCCTGCTCGCCCGGGACCCGCGGGCCAGGGTC 1105

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||   ||| |||    | |  |||||   |||      | | ||||||  |||||||| |
```

```
Db   1106 GCCGTCGAGACCCTGATCACCACCGGTCAGGTCCATATCGCCGGCGAGGTCACCACCACC 1165

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          || |  |    |||||   |  |    ||  ||||||| |     || ||||| |
Db   1166 GCGTACGCGCCCATCGCCCAACTGGTGCGCGACACCGTCCTGTCCATCGGCTACGACTCC 1225

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          || | || ||||||  ||||    ||| || | |||   |  | | ||||||   |||||
Db   1226 TCGGCCAAGGGCTTCGACGGCGCCTCGTGCGGGGTGTCGGTCTCCATCGGCGCGCAGTCC 1285

Qy    301 CCTGACATCAACCAGGGCGTTGAC---------------------------------- 324
          || ||||||  || ||| || |||
Db   1286 CCGGACATCGCCCGGGGTGTGGACACCGCGTACGAGCGGCGGGGCGGGGGCACGGCCCCG 1345

Qy    325 ---------CGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGC 375
               ||| ||| | |   ' |||||||||| |||||||| |||||||| |||
Db   1346 GGCGGACCGGGTGACGAGCTGGACCGGCAGGGCGCGGGCGACCAGGGCCTGATGTTCGGC 1405

Qy    376 TACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTG 435
          |||||     | || |||  || |||||||||     || |||| |   || ||||| ||
Db   1406 TACGCCTGCGACGAGACCCCCGAGCTGATGCCGCTGCCGATCAACCTGGCGCACCGGCTC 1465

Qy    436 GTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCG 495
            ||| | | | || ||||| || ||||||||  | || |  || ||||| ||||
Db   1466 TCCCGGCGGCTGTCGGAGGTGCGGAAGAACGGCACGATCCCCTACCTCCGCCCCGACGGC 1525

Qy    496 AAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTG 555
          || | |||||| ||  |  |||| |||| || |||    ||  |  ||||  | ||||||
Db   1526 AAGACCCAGGTCACCATCGAGTACGACGGCGACAAGGCGGTCCGCCTCGACACGGTCGTG 1585

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615
           | ||| | ||||||  |    |  ||||||| | | |||||  |  |   | |  ||
Db   1586 GTCTCCTCCCAGCACGCCTCCGGCATCGACCTGGACTCGCTACTGGCGCCCGACATCCGC 1645

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCA-- 673
            |   ||  |       | | ||| | |  |          | || |||  |   ||
Db   1646 CGGCATGTCGTGGAGCCCGTCCTCGCCGGACTGGCCGAGGACGGCATCAAGCTGGACACC 1705

Qy    674 -------------------TCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714
                              |||| || ||||| || |||  ||||| ||||| |||||
Db   1706 GCGGGCTACCGGCTGCTGGTCAATCCCACCGGCCGGTTCGAGATCGGCGGCCCGATGGGC 1765

Qy    715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774
          |||   || ||||| || || || || ||| | || ||||||||||||||||| || ||
Db   1766 GACGCGGGCCTGACCGGCCGCAAGATCATCATCGACACCTACGGCGGCATGGCCCGGCAT 1825

Qy    775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834
          || |||||||| |||| || || || || || |||||||||| || || ||||||  |
Db   1826 GGCGGCGGTGCCTTCTCCGGCAAGGACCCGTCCAAGGTGGACCGTTCGGCGGCGTACGCA 1885

Qy    835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894
            || |  || || || ||| |||| || || || |||||   || |||||  | |||
Db   1886 ATGCGCTGGGTGGCCAAGAACGTCGTGGCGGCGGGGCTGGCCTCGCGGTGTGAGGTGCAG 1945

Qy    895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 954
          ||  |||||||| ||||||  ||| || |||  |     |  | || || || ||||| ||
Db   1946 GTCGCCTACGCCATCGGCAAGGCCGAGCCGGTCGGTCTGTTCGTGGAGACCTTCGGCACC 2005

Qy    955 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCA 1014
          |  |  |||       || |  |    | |  | |  | | ||| ||||||||| |||||
Db   2006 GCCACCGTGGACGTGGAGCGGATCGAGCAGGCGATCGGCGAGGTCTTCGACCTCCGCCCG 2065

Qy   1015 TACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1074
            ||    | || | |   ||||| ||||| | ||||||||||     | ||||| || |||
Db   2066 GCCGCGATCATCCGGGATCTGGACCTGCTCCGCCCGATCTACGCCAAGACCGCCGCCTAC 2125

Qy   1075 GGTCACTTTGGTCGTGAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTG 1128
          || ||||| || || ||||      | ||| | ||||||   ||||||    |  ||| |
Db   2126 GGCCACTTCGGCCGCGAACTGCCCGAGTTCACCTGGGAACGCACCGACCGCACCGAGCAG 2185

Qy   1129 CTGCGCGATGCTGCCGGTCT 1148
          ||   ||  || ||||| ||
Db   2186 CTCATCGCCGCGGCCGGCCT 2205
```

```
RESULT 3
US-08-955-957A-6
; Sequence 6, Application US/08955957A
; Patent No. 6312920
;  GENERAL INFORMATION:
;    APPLICANT:  DeHoff, Bradley S.
;    APPLICANT:  Rosteck Jr., Paul R.
;    TITLE OF INVENTION:  SAM Operon
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Eli Lilly and Company
;      STREET:  Lilly Corporate Center
;      CITY:  Indianapolis
;      STATE:  Indiana
;      COUNTRY:  U.S.
;      ZIP:  46285
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/955,957A
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Webster, Thomas D.
;      REGISTRATION NUMBER:  39,872
;      REFERENCE/DOCKET NUMBER:  P-10162
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  317-276-3334
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  4848 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  mRNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
US-08-955-957A-6

  Query Match             32.4%;  Score 374;  DB 4;  Length 4848;
  Best Local Similarity   49.7%;  Pred. No. 4.2e-101;
  Matches  606;  Conservative  122;  Mismatches  420;  Indels   72;  Gaps    3;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
            |:| ||   | ||: :: || :| |||:|||:  | || || || || ||||| |: ||
Db    986 AUGUCACGUCGCCUGUUCACCUCGGAGUCCGUGACCGAGGGCCACCCCGACAAGAUCGCC 1045

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||  |:    ||  |||:  : || ||  : |:||  | ||| |||   ||  | |:
Db   1046 GACCGGAUCAGCGACACCGUCCUCGAUGCCCUGCUCGCCCGGGACCCGCGGGCCAGGGUC 1105

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||   ||| |||    : |  |||||   ||:     : | ||||||  :||||||| |
Db   1106 GCCGUCGAGACCCUGAUCACCACCGGUCAGGUCCAUAUCGCCGGCGAGGUCACCACCACC 1165

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           || :  |    ||:||   |  :    ||  |||||||: |     |: |||:| |
Db   1166 GCGUACGCGCCCAUCGCCCAACUGGUGCGCGACACCGUCCUGUCCAUCGGCUACGACUCC 1225

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           :| | || ||||:: ||||    |:| :| | ||:    |  | |:|||   |||:|
Db   1226 UCGGCCAAGGGCUUCGACGGCGCCUCGUGCGGGGUGUCGGUCUCCAUCGGCGCGCAGUCC 1285

Qy    301 CCTGACATCAACCAGGGCGTTGAC---------------------------------- 324
           || ||||:|  || ||| |: |||
Db   1286 CCGGACAUCGCCCGGGGUGUGGACACCGCGUACGAGCGGCGGGGCGGGGGCACGGCCCCG 1345

Qy    325 ---------CGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGC 375
```

```
Db   1346 GGCGGACCGGGUGACGAGCUGGACCGGCAGGGCGCGGGCGACCAGGGCCUGAUGUUCGGC 1405

Qy    376 TACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTG 435

Db   1406 UACGCCUGCGACGAGACCCCCGAGCUGAUGCCGCUGCCGAUCAACCUGGCGCACCGGCUC 1465

Qy    436 GTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCG 495

Db   1466 UCCCGGCGGCUGUCGGAGGUGCGGAAGAACGGCACGAUCCCCUACCUCCGCCCCGACGGC 1525

Qy    496 AAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTG 555

Db   1526 AAGACCCAGGUCACCAUCGAGUACGACGGCGACAAGGCGGUCCGCCUCGACACGGUCGUG 1585

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615

Db   1586 GUCUCCUCCCAGCACGCCUCCGGCAUCGACCUGGACUCGCUACUGGCGCCCGACAUCCGC 1645

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCA-- 673

Db   1646 CGGCAUGUCGUGGAGCCCGUCCUCGCCCGGACUGGCCGAGGACGGCAUCAAGCUGGACACC 1705

Qy    674 -------------------TCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714

Db   1706 GCGGGCUACCGGCUGCUGGUCAAUCCCACCGGCCGGUUCGAGAUCGGCGGCCCGAUGGGC 1765

Qy    715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774

Db   1766 GACGCGGGCCUGACCGGCCGCAAGAUCAUCAUCGACACCUACGGCGGCAUGGCCCGGCAU 1825

Qy    775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834

Db   1826 GGCGGCGGUGCCUUCUCCGGCAAGGACCCGUCCAAGGUGGACCGUUCGGCGGCGUACGCA 1885

Qy    835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894

Db   1886 AUGCGCUGGGUGGCCAAGAACGUCGUGGCGGCGGGGCUGGCCUCGCGGUGUGAGGUGCAG 1945

Qy    895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 954

Db   1946 GUCGCCUACGCCAUCGGCAAGGCCGAGCCGGUCGGUCUGUUCGUGGAGACCUUCGGCACC 2005

Qy    955 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCA 1014

Db   2006 GCCACCGUGGACGUGGAGCGGAUCGAGCAGGCGAUCGGCGAGGUCUUCGACCUCCGCCCG 2065

Qy   1015 TACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1074

Db   2066 GCCGCGAUCAUCCGGGAUCUGGACCUGCUCCGCCCGAUCUACGCCAAGACCGCCGCCUAC 2125

Qy   1075 GGTCACTTTGGTCGTGAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTG 1128

Db   2126 GGCCACUUCGGCCGCGAACUGCCCGAGUUCACCUGGGAACGCACCGACCGCACCGAGCAG 2185

Qy   1129 CTGCGCGATGCTGCCGGTCT 1148

Db   2186 CUCAUCGCCGCGGCCGGCCU 2205


RESULT 4
US-08-403-852D-4
; Sequence 4, Application US/08403852D
; Patent No. 5891695
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852D
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1208 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1208
US-08-403-852D-4

  Query Match             32.0%;  Score 369.8;  DB 2;  Length 1208;
  Best Local Similarity   59.4%;  Pred. No. 4e-100;
  Matches  716;  Conservative    0;  Mismatches  432;  Indels   57;  Gaps    3;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Db       1 ATGTCCCGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 60

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Db      61 GACCAGATCAGTGACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTC 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Db     121 GCGGTCGAGACCCTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAG 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Db     181 GCGTACGCGCCCATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCG 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Db     241 TCCGCCAAGGGCTTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCC 300
```

```
Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGC----------------------------C 330
          || |||||| |||||||||| ||| | ||                            |
Db    301 CCGGACATCGCCCAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGAC 360

Qy    331 GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAA 390
          || |      | |||||||| || |||||||| |||||||| |||||||| ||  | ||
Db    361 GAGCTCGACCAGCAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAG 420

Qy    391 ACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCT 450
          ||| | |||||||||    || |||       || ||||| ||    |  || | |   |
Db    421 ACCCCCTCGCTGATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACC 480

Qy    451 GAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACT 510
          || || || ||  |||||||  | || |  |||||||| ||||  || | |||||| ||
Db    481 GAGGTCCGCAAGGACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACC 540

Qy    511 TTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCAC 570
           |  |||| | | | ||       ||  | | ||  | ||||  | ||| | ||||||
Db    541 ATCGAGTACCAGGGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCAC 600

Qy    571 TCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCA 630
           | |  || |||||||     || ||||      | |  ||    |||  | ||
Db    601 GCCGCCGACATCGACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAG 660

Qy    631 ATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCT---------------------- 670
              | | ||| | |             | || |||  ||
Db    661 CACGTCCTCGCCGCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTG 720

Qy    671 -TCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACT 729
           |  ||||||||||||| |||||||  ||||| ||||| ||||| ||| |||| |||||
Db    721 CTGGTCAACCCGACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACC 780

Qy    730 GGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTC 789
          || || || || ||| | || || ||||||||||||||| || ||||| || || || |||
Db    781 GGCCGCAAGATCATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTC 840

Qy    790 TCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCG 849
          || || || || || || || || ||||||||||| || |||||    || |  |||||
Db    841 TCCGGCAAGGACCCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCC 900

Qy    850 AAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATC 909
          || ||| |||| || || ||||| |||   || || ||  | |||||  ||||||| |||
Db    901 AAGAACGTCGTCGCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATC 960

Qy    910 GGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCT 969
          |||  ||| || |||  |  | |  | || || || ||||| || |  |  ||  |
Db    961 GGCAAGGCCGAGCCGGTCGGCCTGTTCGTCGAGACGTTCGGCACCGGCACCGTCGCCCAG 1020

Qy    970 GAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAG 1029
          || |   |      |    |    ||| ||||||||||||||||    |   | || |
Db   1021 GAGCGCATCGAGAAGGCCATCACCGAGGTCTTCGACCTGCGCCCCGCGGCCATCATCCGC 1080

Qy   1030 ATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGT 1089
             || ||||||| || ||||||   |  ||||| || ||||| ||||| || ||
Db   1081 GACCTCGACCTGCTGCGGCCCATCTACGCCGCCACCGCCGCCTACGGCCACTTCGGCCGC 1140

Qy   1090 GAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCTGCC 1143
          ||||      | ||| | |||||    ||||||   || || | |||    |  || |||
Db   1141 GAACTGCCCGACTTCACCTGGGAGCGGACCGACCGCGCCCACCGGCTCAAGGCCGCGGCC 1200

Qy   1144 GGTCT 1148
          |||||
Db   1201 GGTCT 1205


RESULT 5
US-08-510-646B-4
; Sequence 4, Application US/08510646B
; Patent No. 6077699
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  45
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/510,646B
;      FILING DATE:  03-AUG-1995
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 08/403,852
;      FILING DATE:  10-MAY-1995
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-01000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1208 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1208
US-08-510-646B-4

  Query Match             32.0%;  Score 369.8;  DB 3;  Length 1208;
  Best Local Similarity   59.4%;  Pred. No. 4e-100;
  Matches  716;  Conservative    0;  Mismatches  432;  Indels   57;  Gaps    3;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||| |    | ||| || || || |||||||  | || || || || ||||| || ||
Db      1 ATGTCCCGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 60

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          ||||| ||   |||  ||||  | |||||  | ||     |||| ||     |||||||
Db     61 GACCAGATCAGTGACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTC 120

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||   ||| |||    | |  ||||||  |||     | | ||||||  |||||||||
Db    121 GCGGTCGAGACCCTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAG 180
```

```
Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          || |  |    |||||   |  |   |||  ||||  | |  |  || ||||| |
Db    181 GCGTACGCGCCCATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCG 240

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          |||| || |||||| ||||    ||||||  |  ||      | | ||||||   |||||
Db    241 TCCGCCAAGGGCTTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCC 300

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGC----------------------------C 330
          || ||||||  |||||||||| ||| | ||                            |
Db    301 CCGGACATCGCCCAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGAC 360

Qy    331 GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAA 390
          || |      | |||||||| || |||||||| |||||||| |||||||| ||  | ||
Db    361 GAGCTCGACCAGCAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAG 420

Qy    391 ACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCT 450
          |||  |  |||||||||    || |||       || ||||| ||    |  || |   |
Db    421 ACCCCCTCGCTGATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACC 480

Qy    451 GAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAACCCAGGTGACT 510
          || || || ||  |||||||  | || |  |||||||| |||  || | |||||| ||
Db    481 GAGGTCCGCAAGGACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACC 540

Qy    511 TTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCAC 570
          |  ||||  | | | ||      ||  |  | ||  | |||||  | ||| | ||||||
Db    541 ATCGAGTACCAGGGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCAC 600

Qy    571 TCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCA 630
          | |  || |||||||     || ||||       | |  ||    |||  | ||
Db    601 GCCGCCGACATCGACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAG 660

Qy    631 ATTCTGCCCCGCTGAATGGCTGACTTCTGCCACCAAATTCT------------------- 670
          |  | |||  | |             | || |||  ||
Db    661 CACGTCCTCGCCGCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTG 720

Qy    671 -TCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACT 729
           |  ||||||||||||| |||||||  ||||| ||||| ||||| |||  ||| |||||
Db    721 CTGGTCAACCCGACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACC 780

Qy    730 GGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTC 789
          || || || || ||| | || || |||||||||||||| || ||||| || || || |||
Db    781 GGCCGCAAGATCATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTC 840

Qy    790 TCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCG 849
          || || || || || || || || ||||||||||| || |||||    || |  |||||
Db    841 TCCGGCAAGGACCCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCC 900

Qy    850 AAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATC 909
          || ||| |||| || || |||| |||   || || ||  | |||||  |||||| |||
Db    901 AAGAACGTCGTCGCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATC 960

Qy    910 GGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCT 969
          |||  ||| || |||  |  | |  | || || || |||| || |  |  ||  |
Db    961 GGCAAGGCCGAGCCGGTCGGCCTGTTCGTCGAGACGTTCGGCACCGGCACCGTCGCCCAG 1020

Qy    970 GAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAG 1029
          || |   |      |    |    ||| ||||||||||||||||    |   | || |
Db   1021 GAGCGCATCGAGAAGGCCATCACCGAGGTCTTCGACCTGCGCCCCGCGGCCATCATCCGC 1080

Qy   1030 ATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGT 1089
             || || |||||||  || ||||||  |  ||||| || ||||| ||||| || ||
Db   1081 GACCTCGACCTGCTGCGGCCCATCTACGCCGCCACCGCCGCCTACGGCCACTTCGGCCGC 1140

Qy   1090 GAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCTGCC 1143
          ||||      | ||| | |||||    ||||||   || || | |||     |  || |||
Db   1141 GAACTGCCCGACTTCACCTGGGAGCGGACCGACCGCGCCCACCGGCTCAAGGCCGCGGCC 1200

Qy   1144 GGTCT 1148
          |||||
Db   1201 GGTCT 1205
```

```
RESULT 6
US-09-231-818-4
; Sequence 4, Application US/09231818
; Patent No. 6171846
;  GENERAL INFORMATION:
;    APPLICANT:  Blanc, Veronique
;    APPLICANT:  Blanche, Francis
;    APPLICANT:  Crouzet, Joel
;    APPLICANT:  Jacques, Nathalie
;    APPLICANT:  Lacroix, Patricia
;    APPLICANT:  Thibaut, Denis
;    APPLICANT:  Zagorec, Monique
;    APPLICANT:  Debussche, Laurent
;    APPLICANT:  De Crecy-Lagard, Valerie
;    TITLE OF INVENTION:  Polypeptides Involved In The
;    TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;    TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;    NUMBER OF SEQUENCES:  43
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;      STREET:  1300 I Street, N.W., Suite 700
;      CITY:  Washington
;      STATE:  D.C.
;      COUNTRY:  USA
;      ZIP:  20005-3315
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/231,818
;      FILING DATE:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/403,852
;      FILING DATE:  10-MAY-1995
;      APPLICATION NUMBER:  PCT/FR 93/00923
;      FILING DATE:  25-SEP-1993
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  FR 92/11441
;      FILING DATE:  25-SEP-1992
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Meyers, Kenneth J.
;      REGISTRATION NUMBER:  25,146
;      REFERENCE/DOCKET NUMBER:  03806.0054-00000
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (202) 408-4000
;      TELEFAX:  (202) 408-4400
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1208 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..1208
US-09-231-818-4

  Query Match            32.0%;  Score 369.8;  DB 4;  Length 1208;
  Best Local Similarity  59.4%;  Pred. No. 4e-100;
  Matches  716;  Conservative    0;  Mismatches  432;  Indels   57;  Gaps    3;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
             ||| |    | ||| || || || ||||||||  | || || || || ||||| || ||
Db       1 ATGTCCCGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 60
```

```
Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||| ||   |||  ||||  | |||||  | ||     |||| ||     |||||||
Db      61 GACCAGATCAGTGACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTC 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||   ||| |||    | |  ||||||  |||     | | ||||||  |||||||||
Db     121 GCGGTCGAGACCCTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAG 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           || |  |    |||||   |  |   |||  ||||  | |  |  || ||||| |
Db     181 GCGTACGCGCCCATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCG 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           |||| || |||||| ||||    ||||||  |  ||       | | ||||||   |||||
Db     241 TCCGCCAAGGGCTTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCC 300

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGC----------------------------C 330
           || ||||||  |||||||| ||| | ||                            |
Db     301 CCGGACATCGCCCAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGAC 360

Qy     331 GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAA 390
           || |     | |||||||| || |||||||| |||||||| |||||||| ||   | ||
Db     361 GAGCTCGACCAGCAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAG 420

Qy     391 ACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCT 450
           |||  |  |||||||||    || |||      || ||||| ||    |  || |   |
Db     421 ACCCCCTCGCTGATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACC 480

Qy     451 GAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACT 510
           || || || ||  |||||||  | || |  |||||||| ||||  || | |||||| ||
Db     481 GAGGTCCGCAAGGACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACC 540

Qy     511 TTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCAC 570
           |  ||||  | | | ||      ||  |  | ||  | |||||  | ||| | ||||||
Db     541 ATCGAGTACCAGGGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCAC 600

Qy     571 TCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCA 630
           | |  || |||||||     || ||||      | |  ||    |||  | ||
Db     601 GCCGCCGACATCGACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAG 660

Qy     631 ATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCT---------------------- 670
           | | ||| | |            | || |||  ||
Db     661 CACGTCCTCGCCGCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTG 720

Qy     671 -TCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACT 729
            |  |||||||||||||| |||||||  ||||| ||||| ||||| ||| ||| ||||
Db     721 CTGGTCAACCCGACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACC 780

Qy     730 GGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTC 789
           || || || || ||| | || || |||||||||||||| || ||||| || || || |||
Db     781 GGCCGCAAGATCATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTC 840

Qy     790 TCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCG 849
           || || || || || || || || ||||||||||| || |||||    || |  |||||
Db     841 TCCGGCAAGGACCCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCC 900

Qy     850 AAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATC 909
           || ||| |||| || || ||||| |||   || || ||  | |||||  ||||||| |||
Db     901 AAGAACGTCGTCGCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATC 960

Qy     910 GGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCT 969
           |||  ||| || |||  |  | |  | || || || ||||| || |  |  ||  |
Db     961 GGCAAGGCCGAGCCGGTCGGCCTGTTCGTCGAGACGTTCGGCACCGGCACCGTCGCCCAG 1020

Qy     970 GAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAG 1029
           || |   |      |    |    ||| ||||||||||||||||    |   | || |
Db    1021 GAGCGCATCGAGAAGGCCATCACCGAGGTCTTCGACCTGCGCCCCGCGGCCATCATCCGC 1080

Qy    1030 ATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGT 1089
             || || |||||||  || ||||||   |  ||||| || ||||| ||||| || ||
Db    1081 GACCTCGACCTGCTGCGGCCCATCTACGCCGCCACCGCCGCCTACGGCCACTTCGGCCGC 1140

Qy    1090 GAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCTGCC 1143
```

```
           ||||      | ||| | |||||    ||||||   || || | |||    |  || |||
Db    1141 GAACTGCCCGACTTCACCTGGGAGCGGACCGACCGCGCCCACCGGCTCAAGGCCGCGGCC 1200

Qy    1144 GGTCT 1148
           |||||
Db    1201 GGTCT 1205
```

RESULT 7
US-08-403-852D-1
; Sequence 1, Application US/08403852D
; Patent No. 5891695
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/403,852D
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5392 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
US-08-403-852D-1

Query Match 32.0%; Score 369.8; DB 2; Length 5392;
Best Local Similarity 59.4%; Pred. No. 7.8e-100;
Matches 716; Conservative 0; Mismatches 432; Indels 57; Gaps 3;

```
Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           |||  |    | ||| || || || ||||||||  | || || || || ||||| || ||
```

```
Db   3558 ATGTCCCGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 3617

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Db   3618 GACCAGATCAGTGACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTC 3677

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Db   3678 GCGGTCGAGACCCTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAG 3737

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Db   3738 GCGTACGCGCCCATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCG 3797

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Db   3798 TCCGCCAAGGGCTTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCC 3857

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGC----------------------------C 330

Db   3858 CCGGACATCGCCCAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGAC 3917

Qy    331 GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAA 390

Db   3918 GAGCTCGACCAGCAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAG 3977

Qy    391 ACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCT 450

Db   3978 ACCCCCTCGCTGATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACC 4037

Qy    451 GAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACT 510

Db   4038 GAGGTCCGCAAGGACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACC 4097

Qy    511 TTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCAC 570

Db   4098 ATCGAGTACCAGGGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCAC 4157

Qy    571 TCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCA 630

Db   4158 GCCGCCGACATCGACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAG 4217

Qy    631 ATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCT---------------------- 670

Db   4218 CACGTCCTCGCCGCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTG 4277

Qy    671 -TCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACT 729

Db   4278 CTGGTCAACCCGACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACC 4337

Qy    730 GGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTC 789

Db   4338 GGCCGCAAGATCATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTC 4397

Qy    790 TCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCG 849

Db   4398 TCCGGCAAGGACCCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCC 4457

Qy    850 AAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATC 909

Db   4458 AAGAACGTCGTCGCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATC 4517

Qy    910 GGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCT 969

Db   4518 GGCAAGGCCGAGCCGGTCGGCCTGTTCGTCGAGACGTTCGGCACCGGCACCGTCGCCCAG 4577

Qy    970 GAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAG 1029

Db   4578 GAGCGCATCGAGAAGGCCATCACCGAGGTCTTCGACCTGCGCCCCGCGGCCATCATCCGC 4637

Qy   1030 ATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGT 1089

Db   4638 GACCTCGACCTGCTGCGGCCCATCTACGCCGCCACCGCCGCCTACGGCCACTTCGGCCGC 4697

Qy   1090 GAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCTGCC 1143

Db   4698 GAACTGCCCGACTTCACCTGGGAGCGGACCGACCGCGCCCACCGGCTCAAGGCCGCGGCC 4757

Qy   1144 GGTCT 1148

Db   4758 GGTCT 4762
```

RESULT 8
US-08-510-646B-1
; Sequence 1, Application US/08510646B
; Patent No. 6077699
; GENERAL INFORMATION:
; APPLICANT: Blanc, Veronique
; APPLICANT: Blanche, Francis
; APPLICANT: Crouzet, Joel
; APPLICANT: Jacques, Nathalie
; APPLICANT: Lacroix, Patricia
; APPLICANT: Thibaut, Denis
; APPLICANT: Zagorec, Monique
; APPLICANT: Debussche, Laurent
; APPLICANT: De Crecy-Lagard, Valerie
; TITLE OF INVENTION: Polypeptides Involved In The
; TITLE OF INVENTION: Biosynthesis Of Streptogramins, Nucleotide Sequences
; TITLE OF INVENTION: Coding For These Polypeptides And Their Use
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett & Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/510,646B
; FILING DATE: 03-AUG-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,852
; FILING DATE: 10-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR 93/00923
; FILING DATE: 25-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 92/11441
; FILING DATE: 25-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Meyers, Kenneth J.
; REGISTRATION NUMBER: 25,146
; REFERENCE/DOCKET NUMBER: 03806.0054-01000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5392 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: S.pristinaespiralis
US-08-510-646B-1

```
  Query Match             32.0%;  Score 369.8;  DB 3;  Length 5392;
  Best Local Similarity   59.4%;  Pred. No. 7.8e-100;
  Matches  716;  Conservative    0;  Mismatches  432;  Indels   57;  Gaps    3;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
Db    3558 ATGTCCCGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 3617

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
Db    3618 GACCAGATCAGTGACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTC 3677

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
Db    3678 GCGGTCGAGACCCTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAG 3737

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
Db    3738 GCGTACGCGCCCATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCG 3797

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
Db    3798 TCCGCCAAGGGCTTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCC 3857

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGC----------------------------C 330
Db    3858 CCGGACATCGCCCAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGAC 3917

Qy     331 GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAA 390
Db    3918 GAGCTCGACCAGCAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAG 3977

Qy     391 ACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCT 450
Db    3978 ACCCCCTCGCTGATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACC 4037

Qy     451 GAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACT 510
Db    4038 GAGGTCCGCAAGGACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACC 4097

Qy     511 TTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCAC 570
Db    4098 ATCGAGTACCAGGGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCAC 4157

Qy     571 TCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCA 630
Db    4158 GCCGCCGACATCGACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAG 4217

Qy     631 ATTCTGCCCCGCTGAATGGCTGACTTCTGCCACCAAATTCT---------------------- 670
Db    4218 CACGTCCTCGCCGCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTG 4277

Qy     671 -TCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACT 729
Db    4278 CTGGTCAACCCGACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACC 4337

Qy     730 GGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTC 789
Db    4338 GGCCGCAAGATCATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTC 4397

Qy     790 TCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCG 849
Db    4398 TCCGGCAAGGACCCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCC 4457

Qy     850 AAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATC 909
Db    4458 AAGAACGTCGTCGCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATC 4517

Qy     910 GGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCT 969
Db    4518 GGCAAGGCCGAGCCGGTCGGCCTGTTCGTCGAGACGTTCGGCACCGGCACCGTCGCCCAG 4577

Qy     970 GAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAG 1029
```

```
Db    4578 GAGCGCATCGAGAAGGCCATCACCGAGGTCTTCGACCTGCGCCCCGCGGCCATCATCCGC 4637

Qy    1030 ATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGT 1089
Db    4638 GACCTCGACCTGCTGCGGCCCATCTACGCCGCCACCGCCGCCTACGGCCACTTCGGCCGC 4697

Qy    1090 GAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCTGCC 1143
Db    4698 GAACTGCCCGACTTCACCTGGGAGCGGACCGACCGCGCCCACCGGCTCAAGGCCGCGGCC 4757

Qy    1144 GGTCT 1148
Db    4758 GGTCT 4762


RESULT 9
US-09-231-818-1
; Sequence 1, Application US/09231818
; Patent No. 6171846
; GENERAL INFORMATION:
;   APPLICANT:  Blanc, Veronique
;   APPLICANT:  Blanche, Francis
;   APPLICANT:  Crouzet, Joel
;   APPLICANT:  Jacques, Nathalie
;   APPLICANT:  Lacroix, Patricia
;   APPLICANT:  Thibaut, Denis
;   APPLICANT:  Zagorec, Monique
;   APPLICANT:  Debussche, Laurent
;   APPLICANT:  De Crecy-Lagard, Valerie
;   TITLE OF INVENTION:  Polypeptides Involved In The
;   TITLE OF INVENTION:  Biosynthesis Of Streptogramins, Nucleotide Sequences
;   TITLE OF INVENTION:  Coding For These Polypeptides And Their Use
;   NUMBER OF SEQUENCES:  43
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Finnegan, Henderson, Farabow, Garrett & Dunner
;     STREET:  1300 I Street, N.W., Suite 700
;     CITY:  Washington
;     STATE:  D.C.
;     COUNTRY:  USA
;     ZIP:  20005-3315
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Floppy disk
;     COMPUTER:  IBM PC compatible
;     OPERATING SYSTEM:  PC-DOS/MS-DOS
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/231,818
;     FILING DATE:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/403,852
;     FILING DATE:  10-MAY-1995
;     APPLICATION NUMBER:  PCT/FR 93/00923
;     FILING DATE:  25-SEP-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  FR 92/11441
;     FILING DATE:  25-SEP-1992
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Meyers, Kenneth J.
;     REGISTRATION NUMBER:  25,146
;     REFERENCE/DOCKET NUMBER:  03806.0054-00000
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (202) 408-4000
;     TELEFAX:  (202) 408-4400
; INFORMATION FOR SEQ ID NO:  1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  5392 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
```

```
;    ORIGINAL SOURCE:
;      ORGANISM:  S.pristinaespiralis
US-09-231-818-1

  Query Match             32.0%;  Score 369.8;  DB 4;  Length 5392;
  Best Local Similarity   59.4%;  Pred. No. 7.8e-100;
  Matches  716;  Conservative    0;  Mismatches  432;  Indels   57;  Gaps    3;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
Db    3558 ATGTCCCGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 3617

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
Db    3618 GACCAGATCAGTGACACCGTCCTCGACGCCCTGCTGCGCGAGGACCCCGCCTCACGCGTC 3677

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
Db    3678 GCGGTCGAGACCCTGATCACCACCGGCCAGGTCCACATCGCCGGCGAGGTCACCACCAAG 3737

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
Db    3738 GCGTACGCGCCCATCGCCCAACTGGTCCGCGACACGATCCTGGCCATCGGCTACGACTCG 3797

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
Db    3798 TCCGCCAAGGGCTTCGACGGCGCCTCCTGCGGCGTCTCCGTCTCCATCGGCGCGCAGTCC 3857

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGC----------------------------C 330
Db    3858 CCGGACATCGCCCAGGGCGTCGACAGCGCCTACGAGACCCGCGTCGAGGGCGAGGACGAC 3917

Qy     331 GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAA 390
Db    3918 GAGCTCGACCAGCAGGGCGCCGGCGACCAGGGCCTGATGTTCGGCTACGCCACCGACGAG 3977

Qy     391 ACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCT 450
Db    3978 ACCCCCTCGCTGATGCCGCTGCCCATCGAGCTCGCCCACCGCCTCTCGCGCCGGCTCACC 4037

Qy     451 GAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACT 510
Db    4038 GAGGTCCGCAAGGACGGCACCGTCCCCTACCTGCGCCCCGACGGCAAGACCCAGGTCACC 4097

Qy     511 TTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCAC 570
Db    4098 ATCGAGTACCAGGGCAGCCGCCCGGTGCGCCTGGACACCGTCGTCGTCTCCTCCCAGCAC 4157

Qy     571 TCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCA 630
Db    4158 GCCGCCGACATCGACCTCGGCTCCCTGCTCACCCCCGACATCCGCGAGCACGTCGTCGAG 4217

Qy     631 ATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCT---------------------- 670
Db    4218 CACGTCCTCGCCGCACTCGCCGAGGACGGCATCAAGCTCGAGACGGACAACTACCGCCTG 4277

Qy     671 -TCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACT 729
Db    4278 CTGGTCAACCCGACCGGCCGTTTCGAGATCGGCGGCCCGATGGGCGACGCCGGCCTGACC 4337

Qy     730 GGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTC 789
Db    4338 GGCCGCAAGATCATCATCGACACGTACGGCGGCATGGCCCGCCACGGCGGTGGCGCGTTC 4397

Qy     790 TCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCG 849
Db    4398 TCCGGCAAGGACCCGTCCAAGGTCGACCGTTCCGCCGCGTACGCGATGCGCTGGGTCGCC 4457

Qy     850 AAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATC 909
Db    4458 AAGAACGTCGTCGCCGCGGGCCTCGCCTCCCGCTGCGAGGTCCAGGTCGCCTACGCCATC 4517

Qy     910 GGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCT 969
Db    4518 GGCAAGGCCGAGCCGGTCGGCCTGTTCGTCGAGACGTTCGGCACCGGCACCGTCGCCCAG 4577

Qy     970 GAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCCATACGGTCTGATTCAG 1029
Db    4578 GAGCGCATCGAGAAGGCCATCACCGAGGTCTTCGACCTGCGCCCCGCGGCCATCATCCGC 4637

Qy    1030 ATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGT 1089
Db    4638 GACCTCGACCTGCTGCGGCCCATCTACGCCGCCACCGCCGGCTACGGCCACTTCGGCCGC 4697

Qy    1090 GAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCTGCC 1143
Db    4698 GAACTGCCCGACTTCACCTGGGAGCGGACCGACCGCGCCCACCGGCTCAAGGCCGCGGCC 4757

Qy    1144 GGTCT 1148
Db    4758 GGTCT 4762


RESULT 10
US-09-273-686-1
; Sequence 1, Application US/09273686
; Patent No. 6228625
; GENERAL INFORMATION:
;  APPLICANT: Zalacain, Magdalena
;  APPLICANT: Burnham, Martin K. R.
;  APPLICANT: Biswas, Sanjoy
;  APPLICANT: Brown, James
;  APPLICANT: Ingraham, Karen, A.
;  APPLICANT: Chalker, Alison F.
;  APPLICANT: So, Chi Y.
;  APPLICANT: Holmes, David J.
;  APPLICANT: Van Horn, Stephanie
;  APPLICANT: Warren, Richard L.
;  TITLE OF INVENTION: metK
;  FILE REFERENCE: GM10176
;  CURRENT APPLICATION NUMBER: US/09/273,686
;  CURRENT FILING DATE: 1999-03-22
;  EARLIER APPLICATION NUMBER: 60/106,767
;  EARLIER FILING DATE: 1998-11-03
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 1191
;   TYPE: DNA
;   ORGANISM: Streptococcus pneumoniae
US-09-273-686-1

  Query Match             31.9%;  Score 368.2;  DB 4;  Length 1191;
  Best Local Similarity   60.5%;  Pred. No. 1.2e-99;
  Matches  676;  Conservative    0;  Mismatches  408;  Indels   33;  Gaps    3;

Qy      14 TTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTCTG 73
Db      17 TATTCACGTCTGAATCTGTATCTGAGGGGCATCCGGATAAGATTGCAGACCAAATTTCAG 76

Qy      74 ATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAACCT 133
Db      77 ATGCGATTTTGGATGCTATTTTAGCAAAGGATCCAGAGGCGCACGTTGCTGCTGAAACAG 136

Qy     134 ACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGACA 193
Db     137 CTGTATATACTGGTTCTGTCCACGTTTTTGGTGAAATTTCTACAAATGCCTATGTGGATA 196

Qy     194 TCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATGGGCT 253
Db     197 TTAACCGTGTGGTTCGTGATACCATTGCAGAGATTGGTTATACCAATACAGAATATGGAT 256

Qy     254 TTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAACC 313
Db     257 TTTCTGCTGAGACGGTGGGAGTACACCCATCTTTGGTGGAACAATCTCCTGACATCGCTC 316
```

```
Qy    314 AGGGCGTTGACCGTGCC--------------------------GATCCGCTGGAACAG- 345
Db    317 AAGGTGTTAACGAAGCCTTGGAGGTTCGTGGAAATGCTGATCAAGATCCACTGGACTTGA 376

Qy    346 --GGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGA 403
Db    377 TTGGAGCAGGTGACCAAGGGCTCATGTTTGGATTTGCAGTAGATGAAACAGAAGAGCTTA 436

Qy    404 TGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAA 463
Db    437 TGCCATTGCCAATTGCACTCAGTCATAAATTGATTCGTCGTCTGGCAGAACTTCGTAAGT 496

Qy    464 ACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACG 523
Db    497 CTGGAGAAATTAGCTATCTCCGTCCAGATGCAAAATCACAAGTTACAGTTGAGTACGATG 556

Qy    524 AC---GGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGA 580
Db    557 AAAATGACCGTCCGGTACGTGTAGATACAGTCGTTATTTCTACTCAGCATGATCCAGAGG 616

Qy    581 TCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCG 640
Db    617 CCACTAATGAACAAATCCATCAAGATGTGATTGACAAGGTCATCAAAGAAGTTATTCCAT 676

Qy    641 CTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCG 700
Db    677 CTTCTTATCTTGATGATAAGACAAAATTCTTTATCAATCCGACAGGTCGTTTTGTAATCG 736

Qy    701 GTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCG 760
Db    737 GTGGTCCTCAAGGGGACTCAGGTTTGACTGGTCGTAAGATTATTGTAGATACTTATGGTG 796

Qy    761 GCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTT 820
Db    797 GCTACTCTCGTCATGGTGGTGGTGCCTTCTCTGGTAAAGATGCGACTAAGGTGGATCGTT 856

Qy    821 CCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATC 880
Db    857 CAGCCTCTTATGCGGCTCGCTATATTGCCAAGAATATCGTTGCAGCAGGCCTTGCTAAGA 916

Qy    881 GTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAG 940
Db    917 AGGCAGAAGTGCAGTTGGCCTATGCTATCGGTGTTGCGCAACCTGTTTCTGTTCGTATCG 976

Qy    941 AAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCT 1000
Db    977 ATACTTTCGGTACAGGAACAGTAGCTGAAAGTCAACTTGAAAAAGCGGCTCGTCAAATCT 1036

Qy   1001 TCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAG 1060
Db   1037 TTGACCTTCGCCCTGCAGGGATTATCCAAATGCTGGACCTCAAGCGTCCAATTTACCGTC 1096

Qy   1061 AAACCGCAGCATACGGTCACTTTGGTCGTGAACATTT 1097
Db   1097 AAACATCGGCTTACGGTCACATGGGACGTACAGATAT 1133
```

```
RESULT 11
US-08-961-527-163
; Sequence 163, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  163:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  8494 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-163

  Query Match            31.9%;  Score 368.2;  DB 4;  Length 8494;
  Best Local Similarity  60.5%;  Pred. No. 2.9e-99;
  Matches  676;  Conservative    0;  Mismatches  408;  Indels   33;  Gaps    3;
```

```
Qy     14 TTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTCTG 73
Db   4736 TATTCACGTCTGAATCTGTATCTGAGGGGCATCCGGATAAGATTGCAGACCAAATTTCAG 4795

Qy     74 ATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAACCT 133
Db   4796 ATGCGATTTTGGATGCTATTTTAGCAAAGGATCCAGAGGCGCACGTTGCTGCTGAAACAG 4855

Qy    134 ACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGACA 193
Db   4856 CTGTATATACTGGTTCTGTCCACGTTTTTGGTGAAATTTCTACAAATGCCTATGTGGATA 4915

Qy    194 TCGAAGAGATCACCCGTAACACCGTTCGCCAAATTGGCTATGTGCATTCCGACATGGGCT 253
Db   4916 TTAACCGTGTGGTTCGTGATACCATTGCAGAGATTGGTTATACCAATACAGAATATGGAT 4975

Qy    254 TTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAACC 313
Db   4976 TTTCTGCTGAGACGGTGGGAGTACACCCATCTTTGGTGGAACAATCTCCTGACATCGCTC 5035

Qy    314 AGGGCGTTGACCGTGCC--------------------------GATCCGCTGGAACAG- 345
Db   5036 AAGGTGTTAACGAAGCCTTGGAGGTTCGTGGAAATGCTGATCAAGATCCACTGGACTTGA 5095

Qy    346 --GGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGA 403
Db   5096 TTGGAGCAGGTGACCAAGGGCTCATGTTTGGATTTGCAGTAGATGAAACAGAAGAGCTTA 5155

Qy    404 TGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAA 463
Db   5156 TGCCATTGCCAATTGCACTCAGTCATAAATTGGTTCGTCGTCTGGCAGAACTTCGTAAGT 5215

Qy    464 ACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACG 523
Db   5216 CTGGAGAAATTAGCTATCTCCGTCCAGATGCAAAATCACAAGTTACAGTTGAGTACGATG 5275

Qy    524 AC---GGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGA 580
Db   5276 AAAATGACCGTCCGGTACGTGTAGATACAGTCGTTATTTCTACTCAGCATGATCCAGAGG 5335

Qy    581 TCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCG 640
Db   5336 CCACTAATGAACAAATCCATCAAGATGTGATTGACAAGGTCATCAAAGAAGTTATTCCAT 5395
```

```
Qy     641 CTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCG 700
Db    5396 CTTCTTATCTTGATGATAAGACAAAATTCTTTATCAATCCGACAGGTCGTTTTGTAATCG 5455

Qy     701 GTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCG 760
Db    5456 GTGGTCCTCAAGGGGACTCAGGTTTGACTGGTCGTAAGATTATTGTAGATACTTATGGTG 5515

Qy     761 GCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTT 820
Db    5516 GCTACTCTCGTCATGGTGGTGGTGCCTTCTCTGGTAAAGATGCGACTAAGGTGGATCGTT 5575

Qy     821 CCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATC 880
Db    5576 CAGCCTCTTATGCGGCTCGCTATATTGCCAAGAATATCGTTGCAGCAGACCTTGCTAAGA 5635

Qy     881 GTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAG 940
Db    5636 AGGCAGAAGTGCAGTTGGCCTATGCTATCGGTGTTGCGCAACCTGTTTCTGTTCGTATCG 5695

Qy     941 AAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCT 1000
Db    5696 ATACTTTCGGTACAGGAACAGTAGCTGAAAGTCAACTTGAAAAAGCGGCTCGTCAAATCT 5755

Qy    1001 TCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAG 1060
Db    5756 TTGACCTTCGCCCTGCAGGGATTATCCAAATGCTGGACCTCAAGCGTCCAATTTACCGTC 5815

Qy    1061 AAACCGCAGCATACGGTCACTTTGGTCGTGAACATTT 1097
Db    5815 AAACATCGGCTTACGGTCACATGGGACGTACAGATAT 5852


RESULT 12
US-09-134-001C-2240
; Sequence 2240, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2240
;   LENGTH: 1248
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2240

  Query Match             30.2%;  Score 349.2;  DB 4;  Length 1248;
  Best Local Similarity   58.5%;  Pred. No. 5.3e-94;
  Matches  687;  Conservative    0;  Mismatches  448;  Indels   39;  Gaps    3;

Qy       7 AAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAA 66
Db      61 AAACGATTATTTACCTCAGAATCAGTTACTGAAGGCCATCCTGATAAAATTGCTGACCAA 120

Qy      67 ATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGC 126
Db     121 GTTTCAGATGCCATTCTTGATGAAATATTAAAGGATGATCCCAATGCTCGTGTAGCATGT 180

Qy     127 GAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGG 186
Db     181 GAAACAACAGTAACAACAGGAATGGCACTTATTTCTGGAGAAATCTCAACAACTACTTAT 240

Qy     187 GTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGAC 246
Db     241 GTAGATATTCCTAAAGTGGTGAGAGAAACGATTAAAGATATTGGTTATACTAGAGCAAAA 300

Qy     247 ATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGAC 306
Db     301 TATGGTTATGATTCACAAACAATGGCACTGTTAACTGCGATTGATGAACAGTCTCCTGAT 360

Qy     307 ATCAACCAGGGCGTTGACCGTGC----------------------------CGATCCG 336
Db     361 ATTGCACAGGGTGTAGACAAAGCTCTTGAGTATCGAAATGATATTTCTGAAGAAGAAATT 420

Qy     337 CTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGAC 396
Db     421 GAAGCAACAGGTGCAGGTGACCAAGGATTAATGTTTGGATATGCAACTGATGAAACTGAT 480

Qy     397 GTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTG 456
Db     481 ACGTATATGCCTCTACCTATATTCCTGTCGCATCAACTTGCTAAACGATTGGCTGATGTA 540

Qy     457 CGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAG 516
Db     541 CGAAAAGATGAAATTTTAGATTATCTTCGTCCAGATGGAAAAGTACAGGTGACTGTTGAA 600

Qy     517 TA---TGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCT 573
Db     601 TATGGTGAAGATGACAAACCTAGACGTATTGATACCATTGTAGTTTCTACACAACATGCT 660

Qy     574 GAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATT 633
Db     661 GAAGATGTAGAGTTAGCACAAATTGAAAAAGACATTAAAACGCATGTTATTTACCCAACT 720

Qy     634 CTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTC 693
Db     721 GTAGATAAAGCTTTATTAGATGATGAAACTAAATTTTACATTAACCCGACTGGACGTTTC 780

Qy     694 GTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACC 753
Db     781 GTTATTGGAGGACCTCAAGGAGATGCTGGTTTAACTGGACGTAAAATTATAGTTGATACG 840

Qy     754 TACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTG 813
Db     841 TATGGTGGTTATGCCCGTCATGGTGGAGGTTGTTTTAGTGGTAAAGATCCTACTAAAGTA 900

Qy     814 GACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTG 873
Db     901 GATCGTTCAGCAGCTTATGCAGCAAGATATGTAGCTAAAAATATTGTTGCAGCTGGTTTA 960

Qy     874 GCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATC 933
Db     961 GCTAAACAATGTGAAGTACAACTTGCATATGCAATTGGTGTAGCAGAACCGTTTCCATT 1020

Qy     934 ATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGT 993
Db    1021 TCAATTAATACGTTTGATACTGGAAAGGTTTCAGAAGCACGTTTAGTTGAAGCTGTAAGA 1080

Qy     994 GAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATC 1053
Db    1081 AAGCATTTTGATTTAAGACCAGCAGGTATCATTAAAATGTTAGACTTAAAACAACCGATA 1140

Qy    1054 TACAAAGAAACCGCAGCATACGGTCACTTTGGTCGT------GAACATTTCCCGTGGGAA 1107
Db    1141 TATAGACAAACAGCAGCCGTATGGTCATTTTGGACGTACAGACGTATTGTTACCATGGGAA 1200

Qy    1108 AAAACCGACAAAGCGCAGCTGCTGCGCGATGCTG 1141
Db    1201 AAATTAGATAAAGTCAATGTTTTAAAAGATGCTG 1234


RESULT 13
US-09-318-448-22
; Sequence 22, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
```

```
;  APPLICANT: Johnson, William G.
;  APPLICANT: Stenroos, Edward S.
;  TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
;  FILE REFERENCE: 601-1-057
;  CURRENT APPLICATION NUMBER: US/09/318,448
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
;   LENGTH: 1283
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-448-22

  Query Match           28.8%;  Score 332.2;  DB 4;  Length 1283;
  Best Local Similarity  57.3%;  Pred. No. 5.9e-89;
  Matches  648;  Conservative    0;  Mismatches  468;  Indels   15;  Gaps    2;

Qy      3 GGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGA 62
          ||  | |  |||||| || || ||||| |||   ||||| || || || || ||| |||
Db    110 GGGCACATTCCTTTTCACCTCAGAGTCGGTCGGGGAAGGCCACCCAGATAAGATTTGTGA 169

Qy     63 CCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGC 122
          ||||||   |||||| ||  | || ||   |||  | ||||||||  | ||    || ||
Db    170 CCAAATCAGTGATGCTGTCCTTGATGCCCACCTTCAGCAGGATCCTGATGCCAAAGTAGC 229

Qy    123 TTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGC 182
          ||| |||||    |  ||||| || ||| |  |  |||  || ||||| ||  |||| ||
Db    230 TTGTGAAACTGTTGCTAAAACTGGAATGATCCTTCTTGCTGGGGAAATTACATCCAGAGC 289

Qy    183 CTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTC 242
            || |||  | |  |  |    ||| |  |  |||    | ||||| ||||   ||||
Db    290 TGCTGTTGACTACCAGAAAGTGGTTCGTGAAGCTGTTAAACACATTGGATATGATGATTC 349

Qy    243 CGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCC 302
            ||  || ||||||   ||  | |||   || |||   ||  | |   |||||||  ||
Db    350 TTCCAAAGGTTTTGACTACAAGACTTGTAACGTGCTGGTAGCCTTGGAGCAACAGTCACC 409

Qy    303 TGACATCAACCAGGGCGTTGACCGTGCC---GATCCGCTGGAACAGGGCGCGGGTGACCA 359
           ||  ||    || || ||| | |  || |    ||       ||    || || || |||||
Db    410 AGATATTGCTCAAGGTGTTCATCTTGACAGAAATGAAGAAGACATTGGTGCTGGAGACCA 469

Qy    360 GGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCAC 419
          |||  | ||||||||||| || ||| |||||| || | .|   |||||   | | ||
Db    470 GGGCTTAATGTTTGGCTATGCCACTGATGAAACTGAGGAGTGTATGCCTTTAACCATTGT 529

Qy    420 CTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTG 479
          ||  ||||||   ||           | ||| ||| | ||    || |||||| |||| ||
Db    530 CTTGGCACACAAGCTAAATGCCAAACTGGCAGAACTACGCCGTAATGGCACTTTGCCTTG 589

Qy    480 GCTGCGCCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTG- 538
          | | ||||| ||  | ||||  || || ||| | ||||||    | |    |  | ||
Db    590 GTTACGCCCTGATTCTAAAACTCAAGTTACTGTGCAGTATATGCAGGATCGAGGTGCTGT 649

Qy    539 ----------GTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCA 587
                   |  || |  |  | ||  | ||   ||||||   |||||||  |    |
Db    650 GCTTCCCATCAGAGTCCACACAATTGTTATATCTGTTCAGCATGATGAAGAGGTTTGTCT 709

Qy    588 GAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATG 647
            |     ||   || ||  ||| |||  |  |||||||  || || |||| ||  |||
Db    710 TGATGAAATGAGGGATGCCCTAAAGGAGAAAGTCATCAAAGCAGTTGTGCCTGCGAAATA 769

Qy    648 GCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCC 707
            ||   |   | || |  | |  | |  | || |  ||  | || ||||| ||||| ||
Db    770 CCTTGATGAGGATACAATCTACCACCTACAGCCAAGTGGCAGATTTGTTATTGGTGGGCC 829

Qy    708 AATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGC 767
          ||||||    ||| |||||||| || ||||| || || || || ||||    ||
Db    830 TCAGGGTGATGCTGGTTTGACTGGACGGAAAATCATTGTGGACACTTATGGCGGTTGGGG 889

Qy    768 GCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGC 827
            ||| || || ||||| || || || || |||    | || || |||||||| || ||
```

```
Db    890 TGCTCATGGAGGAGGTGCCTTTTCAGGAAAGGATTATACCAAGGTCGACCGTTCAGCTGC 949

Qy    828 CTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGA 887
            || || || ||||  || || |||  | | |||   |  || |||  |    |   |
Db    950 TTATGCTGCTCGTTGGGTGGCAAAATCCCTTGTTAAAGGAGGTCTGTGCCGGAGGGTTCT 1009

Qy    888 AATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTT 947
            ||||||| || || || || || ||  || | ||    || |||    |      ||
Db   1010 TGTTCAGGTCTCTTATGCTATTGGAGTTTCTCATCCATTATCTATCTCCATTTTCCATTA 1069

Qy    948 CGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCT 1007
            |||||     |   |  |     | | ||      | | ||     ||   ||||| ||
Db   1070 TGGTACCTCTCAGAAGAGTGAGAGAGAGCTATTAGAGATTGTGAAGAAGAATTTCGATCT 1129

Qy    1008 GCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGC 1067
            |||||     |   |  |   |   |||||||||  | | || || ||  |    || ||
Db    1130 CCGCCCTGGGGTCATTGTCAGGGATCTGGATCTGAAGAAGCCAATTTATCAGAGGACTGC 1189

Qy    1068 AGCATACGGTCACTTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAA 1118
           ||| || || ||||||||| | ||    ||||| ||||||    || | ||
Db    1190 AGCCTATGGCCACTTTGGTAGGGACAGCTTCCCATGGGAAGTGCCCAAAAA 1240


RESULT 14
US-09-318-448-20
; Sequence 20, Application US/09318448
; Patent No. 6210950
; GENERAL INFORMATION:
;  APPLICANT: Johnson, William G.
;  APPLICANT: Stenroos, Edward S.
;  TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
;  TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
;  FILE REFERENCE: 601-1-057
;  CURRENT APPLICATION NUMBER: US/09/318,448
;  CURRENT FILING DATE: 1999-05-25
;  NUMBER OF SEQ ID NOS: 46
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 20
;   LENGTH: 3228
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-318-448-20

  Query Match           27.4%;  Score 316;  DB 4;  Length 3228;
  Best Local Similarity  56.8%;  Pred. No. 5.8e-84;
  Matches  630;  Conservative    0;  Mismatches  465;  Indels   15;  Gaps    2;

Qy     12 CCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTC 71
          | | || || || ||||| ||    || || || || || || ||   |||||| ||
Db    125 CATGTTCACATCGGAGTCTGTGGGAGAGGGACACCCGGATAAGATCTGTGACCAGATCAG 184

Qy     72 TGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAAC 131
          |||||| ||  | || ||    ||| | || || || || ||    || || || || ||
Db    185 TGATGCAGTGCTGGATGCCCATCTCAAGCAAGACCCCAATGCCAAGGTGGCCTGTGAGAC 244

Qy    132 CTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGA 191
              || |||||||||||  |  |  | || || |||||  | |  |||  ||| ||
Db    245 AGTGTGCAAGACCGGCATGGTGCTGCTGTGTGGTGAGATCACCTCAATGGCCATGGTGGA 304

Qy    192 CATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATGGG 251
          |  | |   | |      | ||||| |      | || |||||  |  | || | || |||
Db    305 CTACCAGCGGGTGGTGAGGGACACCATCAAGCACATCGGCTACGATGACTCAGCCAAGGG 364

Qy    252 CTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAA 311
          |||||||   ||  | ||    || |||   ||| | |   | || || || || ||
Db    365 CTTTGACTTCAAGACTTGCAACGTGCTGGTGGCTTTGGAGCAGCAATCCCCAGATATTGC 424

Qy    312 CCAGGGCGTTGACCGTGCC---GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGAT 368
          |||| ||||   | |  | |    ||  | |||   ||| || || || ||||| ||||
Db    425 CCAGTGCGTCCATCTGGACAGAAATGAGGAGGATGTGGGGGCAGGAGATCAGGGTTTGAT 484

Qy    369 GTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACA 428
          ||| ||||| || ||  | || || || | |   |||||      |||||| |  ||| ||
```

```
Db    485 GTTCGGCTATGCTACCGACGAGACAGAGGAGTGCATGCCCCTCACCATCATCCTTGCTCA 544

Qy    429 CCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCC 488

Db    545 CAAGCTCAACGCCCGGATGGCAGACCTCAGGCGCTCCGGCCTCCTCCCCTGGCTGCGGCC 604

Qy    489 GGACGCGAAAAGCCAGGTGACTTTTCAGT----------ATGACGACGGCAAAATCGT 536

Db    605 TGACTCTAAGACTCAGGTGACAGTTCAGTACATGCAGGACAATGGCGCAGTCATCCCTGT 664

Qy    537 TGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCT 596

Db    665 GCGCATCCACACCATCGTCATCTCTGTGCAGCACAACGAAGACATCACGCTGGAGGAGAT 724

Qy    597 GCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTC 656

Db    725 GCGCAGGGCCCTGAAGGAGCAAGTCATCAGGGCCGTGGTGCCGGCCAAGTACCTGGACGA 784

Qy    657 TGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGA 716

Db    785 AGACACCGTCTACCACCTGCAGCCCAGTGGGCGGTTTGTCATCGGAGGTCCCCAGGGGGA 844

Qy    717 CTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGG 776

Db    845 TGCGGGTGTCACTGGCCGTAAGATTATTGTGGACACCTATGGCGGCTGGGGGGCTCATGG 904

Qy    777 TGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGC 836

Db    905 TGGTGGGGCCTTCTCTGGGAAGGACTACACCAAGGTAGACCGCTCAGCTGCATATGCTGC 964

Qy    837 ACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGT 896

Db    965 CCGCTGGGTGGCCAAGTCTCTGGTGAAAGCAGGGCTCTGCCGGAGAGTGCTTGTCCAGGT 1024

Qy    897 TTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGA 956

Db   1025 TTCCTATGCCATTGGTGTGGCCGAGCCGCTGTCCATTTCCATCTTCACCTACGGAACCTC 1084

Qy    957 GAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATA 1016

Db   1085 TCAGAAGACAGAGCGAGAGCTGCTGGATGTGGTGCATAAGAACTTCGACCTCCGGCCGGG 1144

Qy   1017 CGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGG 1076

Db   1145 CGTCATTGTCAGGGATTTGGACTTGAAGAAGCCCATCTACCAGAAGACAGCATGCTACGG 1204

Qy   1077 TCACTTTGGTCGTGAACATTTCCCGTGGGA 1106

Db   1205 CCATTTCGGAAGAAGCGAGTTCCCATGGGA 1234


RESULT 15
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
;  APPLICANT: FLEISCHMAN, Robert D.
;  APPLICANT:  WHITE, Owen R.
;  APPLICANT:  FRASER, Claire M.
;  APPLICANT:  VENTER, John C.
;  TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
;  TITLE OF INVENTION:  TUBERCULOSIS
;  FILE REFERENCE: 24366-20007.00
;  CURRENT APPLICATION NUMBER: US/09/103,840A
;  CURRENT FILING DATE:  1998-06-24
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

  Query Match            27.1%;  Score 313.2;  DB 4;  Length 4403765;
  Best Local Similarity  58.2%;  Pred. No. 9.7e-82;
  Matches  658;  Conservative    0;  Mismatches  418;  Indels   54;  Gaps    4;

Qy      13 CTTTTTACGTCCGAGTCCGTCTCTGAAGGCATCCTGACAAAATTGCTGACCAAATTTCT 72

Db 1566680 CTGTTTACCAGTGAGTCGGTGACAGAGGACATCCCGACAAGATCTGTGACGCCATCAGC 1566739

Qy      73 GATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAACC 132

Db 1566740 GACTCGGTTCTGGACGCGCTTCTAGCGGCGGACCCGCGCTCACGTGTCGCGGTCGAGACG 1566799

Qy     133 TACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGAC 192

Db 1566800 CTGGTGACCACCGGGCAGGTGCACGTGGTGGGTGAGGTGACCACCTCGGCTAAGGAGGCG 1566859

Qy     193 ATCGAAGAGATCAC---------CCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCC 243

Db 1566860 TTTGCCGACATCACCAACACGGTCCGCGCACGGATCCTCGAGATCGGCTACGACTCGTCG 1566919

Qy     244 GACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCT 303

Db 1566920 GACAAGGGTTTCGACGGGGCGACCTGCGGGGTGAACATCGGCATCGGCGCACAGTCACCC 1566979

Qy     304 GACATCAACCAGGGCGTTGA--------------------------CCGTGCCGATCCG 336

Db 1566980 GACATCGCCCAGGGGGTCGACACCGCCCACGAGGCCCGGGTCGAGGGCGCGGCCGATCCG 1567039

Qy     337 CTGGA---ACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACC 393

Db 1567040 CTGGACTCCCAGGGCGCCGGTGACCAGGGCCTGATGTTCGGCTACGCGATCAATGCCACC 1567099

Qy     394 GACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAA 453

Db 1567100 CCGGAACTGATGCCACTGCCCATCGCGCTGGCCCACCGACTGTCGCGGCGGCTGACCGAG 1567159

Qy     454 GTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTT 513

Db 1567160 GTCCGCAAGAACGGGGTGCTGCCCTACCTGCGTCCGGATGGCAAGACGCAGGTCACTATC 1567219

Qy     514 CAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCT 573

Db 1567220 GCCTACGAGGACAACGTTCCGGTGCGGCTGGATACCGTGGTCATCTCCACCCAGCACGCG 1567279

Qy     574 GAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATT 633

Db 1567280 GCCGATATCGACCTGGAGAAGACGCTTGATCCCGACATCCGGGAAAAGGTGCTCAACACC 1567339

Qy     634 CTGCCCGCTGA---------------ATGGCTGACTTCTGCCACCAAATTCTTCATCAAC 678

Db 1567340 GTGCTCGACGACCTGGCCCACGAAACCCTGGACGCGTCGACGGTGCGGGTGCTGGTGAAC 1567399

Qy     679 CCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAA 738

Db 1567400 CCGACCGGCAAGTTCGTGCTCGGCGGGCCGATGGGCGATGCCGGGCTCACCGGCCGCAAG 1567459

Qy     739 ATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAA 798

Db 1567460 ATCATCGTCGACACCTACGGCGGCTGGGCCCGCCACGGCGGCGGCGCCTTCTCCGGCAAG 1567519

Qy     799 GATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATC 858

Db 1567520 GATCCGTCCAAGGTGGACCGGTCGGCGGCGTACGCGATGCGCTGGGTGGCCAAGAATGTC 1567579

Qy     859 GTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCT 918

Db 1567580 GTCGCCGCCGGGTTGGCTGAACGGGTCGAGGTGCAGGTGGCCTACGCCATCGGTAAAGCG 1567639

Qy     919 GAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTG 978
```

```
          | |||   |  | |  | || || || |||||||| ||||  |      | |   |  |
Db 1567640 GCACCCGTCGGCCTGTTCGTCGAGACGTTCGGTACCGAGACGGAAGACCCGGTCAAGATC 1567699

Qy    979 ACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGAT 1038
              |    |  | ||| | ||||||||||||||||    |   | || |     ||| |
Db 1567700 GAGAAGGCCATCGGCGAGGTATTCGACCTGCGCCCCGGTGCCATCATCCGCGACCTGAAC 1567759

Qy   1039 CTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCG 1088
       · ||| ||| |||||||||       ||||| || ||||| ||||| || ||
Db 1567760 CTGTTGCGCCCGATCTATGCGCCGACCGCCGCCTACGGGCACTTCGGCCG 1567809


Search completed: April 22, 2003, 14:26:12
Job time : 4798.39 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 12:23:08 ; Search time 147.738 Seconds
(without alignments)
8506.861 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155
Sequence: 1 atggcaaaacacctttttac.....atgctgccggtctgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1155	100.0	1155	10	US-09-815-242-6198	Sequence 6198, Ap
	2	926.8	80.2	1167	10	US-09-815-242-9641	Sequence 9641, Ap
	3	641.4	55.5	1155	10	US-09-815-242-7099	Sequence 7099, Ap
	4	586.2	50.8	1191	10	US-09-815-242-7694	Sequence 7694, Ap
	5	391.4	33.9	1203	10	US-09-974-300-972	Sequence 972, App
	6	375.4	32.5	1191	10	US-09-815-242-4516	Sequence 4516, Ap
	7	375.4	32.5	1296	10	US-09-815-242-8183	Sequence 8183, Ap
	8	375.4	32.5	1296	10	US-09-815-242-8758	Sequence 8758, Ap
c	9	375.4	32.5	10320	7	US-08-781-986A-143	Sequence 143, App
	10	374	32.4	4848	9	US-09-943-702-1	Sequence 1, Appli
	11	374	32.4	4848	9	US-09-943-702-4	Sequence 4, Appli
	12	374	32.4	4848	9	US-09-943-702-6	Sequence 6, Appli
	13	370.6	32.1	1185	10	US-09-815-242-6669	Sequence 6669, Ap
	14	369.8	32.0	1191	10	US-09-815-242-9315	Sequence 9315, Ap
	15	369.2	32.0	1158	10	US-09-815-242-3823	Sequence 3823, Ap
	16	368.2	31.9	1191	10	US-09-815-242-9493	Sequence 9493, Ap
	17	368.2	31.9	1191	10	US-09-805-847-1	Sequence 1, Appli
c	18	356.6	30.9	1377	10	US-09-070-927A-672	Sequence 672, App
	19	331.8	28.7	1221	9	US-09-738-626-1773	Sequence 1773, Ap
	20	331.8	28.7	1344	9	US-09-746-660A-119	Sequence 119, App
	21	314.4	27.2	1487	10	US-09-880-107-1665	Sequence 1665, Ap
	22	309.8	26.8	1185	10	US-09-887-576-784	Sequence 784, App
	23	305.6	26.5	1190	10	US-09-887-576-807	Sequence 807, App
	24	296.8	25.7	1158	10	US-09-815-242-7386	Sequence 7386, Ap
	25	292	25.3	1190	10	US-09-887-576-776	Sequence 776, App
	26	285	24.7	1182	10	US-09-887-576-450	Sequence 450, App
	27	264.8	22.9	1173	9	US-09-938-842A-2445	Sequence 2445, Ap
	28	210.8	18.3	609	10	US-09-974-300-5505	Sequence 5505, Ap
c	29	208	18.0	2774	10	US-09-802-853-5	Sequence 5, Appli
	30	162	14.0	495	10	US-09-887-576-800	Sequence 800, App
	31	155.2	13.4	495	10	US-09-887-576-787	Sequence 787, App
	32	129	11.2	442	10	US-09-770-444-871	Sequence 871, App
	33	120	10.4	366	9	US-10-076-622-278	Sequence 278, App
	34	120	10.4	366	10	US-09-604-287A-278	Sequence 278, App
	35	120	10.4	366	10	US-09-339-338-278	Sequence 278, App
	36	120	10.4	366	12	US-10-007-805-278	Sequence 278, App
	37	119.2	10.3	379	10	US-09-878-574-3276	Sequence 3276, Ap
	38	114.6	9.9	411	10	US-09-878-574-3928	Sequence 3928, Ap
	39	113.4	9.8	383	10	US-09-878-574-1980	Sequence 1980, Ap
	40	110	9.5	413	10	US-09-878-574-3703	Sequence 3703, Ap
	41	105.6	9.1	393	10	US-09-878-574-3731	Sequence 3731, Ap
	42	105.2	9.1	394	10	US-09-878-574-1654	Sequence 1654, Ap
	43	104	9.0	389	10	US-09-878-574-2715	Sequence 2715, Ap
	44	102.2	8.8	386	10	US-09-878-574-1498	Sequence 1498, Ap
	45	102.2	8.8	389	10	US-09-878-574-1447	Sequence 1447, Ap

ALIGNMENTS

RESULT 1
US-09-815-242-6198
; Sequence 6198, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6198
; LENGTH: 1155
; TYPE: DNA
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1155)
US-09-815-242-6198

```
  Query Match           100.0%;  Score 1155;  DB 10;  Length 1155;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1155;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360

Qy     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420

Qy     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480

Qy     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540

Qy     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600

Qy     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660

Qy     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720

Qy     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780

Qy     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840

Qy     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900

Qy     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960

Qy     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020

Qy    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080

Qy    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140

Qy    1141 GCCGGTCTGAAGTAA 1155
           |||||||||||||||
Db    1141 GCCGGTCTGAAGTAA 1155
```

RESULT 2
US-09-815-242-9641
; Sequence 9641, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9641
; LENGTH: 1167
; TYPE: DNA
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1167)
; NAME/KEY: misc_feature
; LOCATION: (1)...(1167)
; OTHER INFORMATION: n = A,T,C or G
US-09-815-242-9641

```
  Query Match           80.2%;  Score 926.8;  DB 10;  Length 1167;
  Best Local Similarity  87.6%;  Pred. No. 1.1e-296;
  Matches 1012;  Conservative    0;  Mismatches  143;  Indels    0;  Gaps    0;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
           |||||||||||||||||||||| ||||||| || ||||||||||||||||||| ||
Db      13 ATGGCAAAACACCTTTTTACGTCTGAGTCCGCATCAGAAGGGCATCCTGACAAAATCGCA 72

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
           |||||||| |||||||||| || ||||| |||||  ||||||||||||||||||||||
Db      73 GACCAAATCTCTGATGCCGTGTTGGACGCTATCCTGCAACAGGATCCGAAAGCACGCGTC 132

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
           || || |||||||||| |||||||||||||||||||||||||| || ||||||||||||
```

```
Db    133 GCCTGTGAAACCTACGTTAAAACCGGCATGGTTTTAGTTGGCGGTGAGATCACCACCAGC 192

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          |||||||| || |||||||||| |||||||| || || ||||||||||||||||||||||
Db    193 GCCTGGGTCGATATCGAAGAGATTACCCGTAATACGGTGCGCGAAATTGGCTATGTGCAT 252

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          |||||||||||||||||||| ||||| || ||||| |||||||| || ||||||||||||
Db    253 TCCGACATGGGCTTTGACGCCAACTCTTGCGCGGTACTGAGCGCAATTGGCAAACAGTCT 312

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          || || |||||||||||||||||| || || |||||||||| |||||||| ||||||
Db    313 CCGGATATCAACCAGGGCGTTGACCGCGCAGACCCGCTGGAACAAGGCGCGGGCGACCAG 372

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          || |||||||||||||||| || |||||||||||||| |||||||| || || ||||||
Db    373 GGCCTGATGTTTGGCTACGCGACAAATGAAACCGACGTACTGATGCCTGCGCCAATCACC 432

Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          || || |||||||||||||||||||||||||||||||||||||||| ||||| ||||| |||
Db    433 TACGCGCACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAATGGCACCCTGCCATGG 492

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          |||| ||||| || |||||||||| ||||| |||||||||||||||||||||||||| |||
Db    493 TTGCGTCCGGATGCAAAAAGCCAGGTCACTTTCCAGTATGACGACGGCAAAATCGTCGGT 552

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          ||||| || || || || || || |||||| | ||||| |||||||| ||||||||||||
Db    553 ATCGACGCCGTGGTTCTCTCTACGCAGCACGCAGAAGATATCGACCAAAAATCGCTGCAA 612

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          |||||||| |||||||||||||||||| || |||||||| |||||||| | | | | ||
Db    613 GAAGCGGTGATGGAAGAGATCATCAAACCCATTCTGCCGTCTGAATGGTTAAATACGTCC 672

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          || || || || |||||||| ||||| ||||| |||||||| ||||| ||||| || ||
Db    673 ACTAAGTTTTTTATCAACCCAACCGGGCGTTTTGTTATCGGCGGCCCGATGGGCGATTGT 732

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          |||||||| |||||||| || ||||||||||||||||||| ||||||||||||| |||
Db    733 GGTCTGACCGGTCGTAAGATCATCGTTGATACCTACGGCGGTATGGCGCGTCACGGCGGC 792

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          || || ||||| |||||||||| || ||||| |||||||| ||||| ||||| || |||
Db    793 GGCGCCTTCTCCGGTAAAGATCCGTCTAAAGTTGACCGTTCTGCAGCTTACGCTGCGCGT 852

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          |||||||||||||||||| || || || ||||| ||||| |||||||| ||||| |||
Db    853 TATGTCGCGAAAAACATCGTGGCGGCAGGTCTGGCGGATCGCTGTGAAATCCAGGTCTCC 912

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          ||||| |||||||| || || ||||| |||||||||| ||||| |||||||||| |||
Db    913 TACGCTATCGGCGTAGCGGAGCCGACGTCCATCATGGTGGAAACGTTCGGTACTGAAAAA 972

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          |||||| |||||||  |||  |||||||| || || ||||||||||||||||  |||||
Db    973 GTGCCTGCTGAACAGTTGATTCTGCTGGTGCGCGAATTCTTCGACCTGCGCCNGTACGGC 1032

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          |||||||||||||||||||||||||||||||||||||||||||||||| || || |||||
Db   1033 TTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACTGCGGCTTACGGCCAC 1092

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          || ||||| ||| | ||||| |||||||||||||||||||||||| ||||||||||||||
Db   1093 TTCGGTCGCGAAAACTTCCCATGGGAAAAAACCGACAAAGCGCAACTGCTGCGCGATGCT 1152

Qy   1141 GCCGGTCTGAAGTAA 1155
          ||||||||||| |||
Db   1153 GCCGGTCTGAAATAA 1167


RESULT 3
US-09-815-242-7099
```

```
; Sequence 7099, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7099
;  LENGTH: 1155
;  TYPE: DNA
;  ORGANISM: Haemophilus influenzae
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(1155)
US-09-815-242-7099

  Query Match           55.5%;  Score 541.4;  DB 10;  Length 1155;
  Best Local Similarity  72.2%;  Pred. No. 4.5e-202;
  Matches  834;  Conservative    0;  Mismatches  321;  Indels    0;  Gaps    0;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
          ||| | |   | | ||||| || || || ||||||||||||||||| || ||||||||
Db      1 ATGTCTAGTTATTTATTTACATCTGAATCAGTCTCTGAAGGGCATCCAGATAAAATTCCC 60

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
          || ||||||||||||||| ||  | || |  ||||  |||| ||||| |||||||| ||
Db     61 GATCAAATTTCTGATGCGGTACTTGATGAAATCCTAAAACAAGATCCAAAAGCACGAGTA 120

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
          ||||| ||||| |||||||||| || ||||  || ||||| || ||||| || ||
Db    121 GCTTGTGAAACTTACGTAAAAACTGGTATGGCATTGGTTGGTGGTGAAATTACAACATCA 180

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
          || ||||| || || ||  |  | ||||||||     || | || ||||| |||| |||
Db    181 GCATGGGTTGATATTGAGAATTTAACCCGTAAAGTGATTTGTGATATTGGTTATGAACAT 240

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
          || || |||||||||| | | | || |||| || || |  || || || ||||| |||
Db    241 TCTGAAATGGGCTTTGATGGTCATTCTTGTGCAGTGCTTAATGCGATTGGTAAACAATCT 300

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAG 360
          | || || || || |||||||| ||||   ||||  | || || |||||| || ||||||
Db    301 GCAGATATTAATCAAGGCGTTGATCGTGAAAATCCATTAGATCAAGGCGCAGGCGACCAA 360

Qy    361 GGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACC 420
          ||| | |||||||| || || || |||||||| || ||  | ||||| ||| | || ||
Db    361 GGTATTATGTTTGGTTATGCCACCAATGAAACGGATGTATTAATGCCTGCAGCCATTACT 420
```

```
Qy    421 TATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGG 480
          |||||||| ||| |  |  |    ||  ||||||||||||||| ||| |   |  | |||
Db    421 TATGCACATCGTTTAATGGAAAAACAATCTGAAGTGCGTAAAAGCGGTAAATTAGCATGG 480

Qy    481 CTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGT 540
          | ||||| || |||||||||| || |||||  | |||||  ||    ||||| ||||||
Db    481 TTACGCCCAGATGCGAAAAGCCAAGTTACTTTAAAATATGAAGATAATAAAATTGTTGGT 540

Qy    541 ATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAA 600
          | ||||| || |||||||| ||||| || ||||||||  |   ||| |||    | ||
Db    541 GTGGATGCGGTTGTGCTTTCTACTCAACATTCTGAAGAAGTTAGCCAAAAAGATTTACAT 600

Qy    601 GAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCC 660
          ||||  || |||||||| || ||||| ||  |  | ||   |||||||||  ||   |
Db    601 GAAGGTGTGATGGAAGAAATTATCAAACCTGTGTTACCAAGTGAATGGCTTTCTAAAGAA 660

Qy    661 ACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGC 720
          || |||||||||| |||||||| ||||||| |||||||||| |||||||| || ||
Db    661 ACAAAATTCTTCATTAACCCGACTGGTCGTTTTGTTATCGGTGGGCCAATGGGCGATTGT 720

Qy    721 GGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGC 780
          ||| | ||||| |||||||| ||||| ||||| ||||| ||     || ||||||||||
Db    721 GGTTTAACTGGCCGTAAAATCATCGTGGATACTTACGGTGGTGCCGCTCGTCACGGTGGT 780

Qy    781 GGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGT 840
          || |||||||| || ||||| ||||| ||||| |||||||| || || || || ||||||
Db    781 GGCGCATTCTCAGGAAAAGAACCATCTAAAGTTGACCGTTCTGCGGCTTATGCTGCACGT 840

Qy    841 TATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCC 900
          |||||||| ||||| || || || || ||  | || |||||||||||||| ||  ||||
Db    841 TATGTCGCAAAAAATATTGTGGCAGCAGGTTTAGCAGATCGTTGTGAAATCCAACTTTCT 900

Qy    901 TACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAA 960
          || || || || |||||||||| || || |||||||| ||||| || || ||||  |||
Db    901 TATGCGATTGGTGTGGCTGAACCAACATCTATCATGGTCGAAACCTTTGGAACTGGAAAA 960

Qy    961 GTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGT 1020
          ||  |   ||||  | |        | || |||||||| || ||  | || || || ||
Db    961 GTTGCGAATGAATTATTAGTGTCATTAGTTCGTGAGTTTTTTGATTTACGTCCTTATGGC 1020

Qy   1021 CTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCAC 1080
          | ||| | ||| | ||| |  | || || |||||    ||||| ||||| || |||||
Db   1021 TTAATTAAAATGTTAGATTTAATCCAACCAATCTATCGTGAAACGGCAGCTTATGGTCAT 1080

Qy   1081 TTTGGTCGTGAACATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTGCGCGATGCT 1140
          || ||||||||||| ||||| |||||||||   ||    ||  |     | ||   |||
Db   1081 TTCGGTCGTGAACAATTCCCTTGGGAAAAAGTGGATCGCGCAGAAGATTTACGTATTGCG 1140

Qy   1141 GCCGGTCTGAAGTAA 1155
          || ||| | || |||
Db   1141 GCAGGTTTAAAATAA 1155


RESULT 4
US-09-815-242-7694
; Sequence 7694, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7694
;   LENGTH: 1191
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1191)
US-09-815-242-7694

  Query Match             50.8%;  Score 586.2;  DB 10;  Length 1191;
  Best Local Similarity   71.2%;  Pred. No. 9.2e-184;
  Matches  774;  Conservative    0;  Mismatches  313;  Indels    0;  Gaps    0;

Qy     12 CCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTC 71
          | |||| || |||||||||| || ||||| ||||| ||||| || || ||||| || ||
Db     15 CGTTTTCACCTCCGAGTCCGTGTCCGAAGGCCATCCGGACAAGATCGCGGACCAGATCTC 74

Qy     72 TGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAAC 131
          || || ||  | ||||| ||| |||   ||||   | | || ||||| || ||||||||
Db     75 CGACGCGGTGCTGGACGCCATCATCGCCAAGGACAAGTACGCTCGCGTGGCATGCGAAAC 134

Qy    132 CTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGA 191
          |   || || |||||| |||   |  | | ||| ||| |||||||  |||||||| ||
Db    135 CCTGGTCAAGACCGGCGTGGCGATCATCGCCGGTGAAGTCACCACCTCCGCCTGGGTCGA 194

Qy    192 CATCGAAGAGATCACCCGTAACACCGTTCGCCGAAATTGGCTATGTGCATTCCGACATGGG 251
          | | |||||| |    || ||   | |   ||| || |||||  |     |||||| | ||
Db    195 CCTGGAAGAGCTGGTGCGCAAGGTCATCATCGACATCGGCTACGACAGCTCCGACGTCGG 254

Qy    252 CTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAA 311
          ||| || | |  | |||| |  || |||| |   |||||||| |||||    ||||||||
Db    255 CTTCGATGGTGCCACCTGCGGCGTGCTGAACATCATCGGCAAGCAGTCGGTGGACATCAA 314

Qy    312 CCAGGGCGTTGACCGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTT 371
          |||||||||  ||||| ||| | |||   || |||||||| || |||||||| ||||||||
Db    315 CCAGGGCGTCGACCGGGCCAAGCCGGAAGACCAGGGCGCCGGCGACCAGGGCCTGATGTT 374

Qy    372 TGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCG 431
          |||||||| |  || || |||||||||||||||| || || |||   |  | |||||
Db    375 CGGCTACGCCAGCAACGAGACCGACGTGCTGATGCCGGCGCCGATCTGTTTCTCCCACCG 434

Qy    432 TCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGA 491
          |||||  |||| |||||| || | ||| ||   ||||   ||||| ||||||||||||||
Db    435 CCTGGTCGAGCGCCAGGCCGAGGCGCGCAAGTCCGGCCTGCTGCCATGGCTGCGCCCCGA 494

Qy    492 CGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGT 551
          ||| ||  |||||||||| |  | || || | |||||  | || || |||||||| ||
Db    495 CGCCAAGTCCCAGGTGACCTGCCGCTACGAAGGCGGCAAGGTGGTCGGCATCGATGCGGT 554

Qy    552 CGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAAT 611
          ||||| ||||| ||||||      ||  ||  | | ||    |||| || |  || ||
Db    555 GGTGCTGTCCACCCAGCACAACCCGGAAGTCTCCTATAACGACCTGCGCGACGGCGTGAT 614

Qy    612 GGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTT 671
          |||   |||||||||||   | ||||| || ||   ||||     | |||| | |||
Db    615 GGAGCTGATCATCAAGCAGGTGCTGCCGGCCGAGCTGCTGCACAAGGACACCCAGTTCCA 674

Qy    672 CATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGG 731
          ||||||||||||||    ||||| ||||| |||||  | || |||||||| ||||| ||
```

```
Db    675 CATCAACCCGACCGGCAACTTCGTGATCGGCGGCCCGGTCGGCGACTGCGGCCTGACCGG 734

Qy    732 TCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTC 791

Db    735 ACGCAAGATCATCGTCGACTCCTACGGCGGCATGGCCCGCCACGGCGGCGGCGCGTTCTC 794

Qy    792 TGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAA 851

Db    795 CGGCAAGGACCCGTCCAAGGTCGACCGCTCCGCCGCCTACGCCGGCCGCTACGTGGCGAA 854

Qy    852 AAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGG 911

Db    855 GAACATCGTCGCCGCCGGCCTCGCCGAGCGTTGCGAGATCCAGGTTTCCTACGCCATCGG 914

Qy    912 CGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGA 971

Db    915 CGTGGCCCAGCCGACCTCGATCTCGATCAACACCTTCGGCACCGGCAAGGTCAGCGACGA 974

Qy    972 ACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGAT 1031

Db    975 GAAGATCGTGCAACTGGTCCGCGAGCACTTCGACCTGCGCCCGTACGCGATCACCAAGAT 1034

Qy   1032 GCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGA 1091

Db   1035 GCTCGACCTGCTCCACCCGATGTACCAGCCGACCGCGGCCTACGGCCACTTCGGTCGTCA 1094

Qy   1092 ACATTTC 1098

Db   1095 TCCGTTC 1101


RESULT 5
US-09-974-300-972
; Sequence 972, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 972
;   LENGTH: 1203
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-972

  Query Match             33.9%;  Score 391.4;  DB 10;  Length 1203;
  Best Local Similarity   60.9%;  Pred. No. 3.6e-119;
  Matches  713;  Conservative    0;  Mismatches  416;  Indels   42;  Gaps    3;

Qy      14 TTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTCTG 73

Db      20 TATTTACATCAGAGTCAGTTACGGAGGGACATCCGGACAAAATTTGCGACCAGATTTCGG 79

Qy      74 ATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAACCT 133

Db      80 ATAGTATATTAGACGAAATTTTAAAGAAAGATCCGAATGCCCGCGTTGCCTGTAAAACAT 139

Qy     134 ACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGACA 193

Db     140 CTGTGACAACCGGATTGGTGCTGGTAAGCGGTGAAATCACGACGTCCACATACGTCGATA 199

Qy     194 TCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATGGGCT 253

Db    200 TTCCGAAAACAGTCCGCGAAACGATTAAAGAAATTGGCTACACACGTGCAAAATACGGTT 259

Qy    254 TTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAACC 313

Db    260 TTGATGCGGAAACTTGCGCGGTCTTAACGTCTATAGATGAACAGTCTCCAGACATTGCAA 319

Qy    314 AGGGCGTTGACC----------------------------------GTGCCGATCCGCTGG 340

Db    320 TGGGAGTCGACCAGGCGCTTGAAGCGCGTGAAGGAGCAATGAGCGACGCTGAAATTGAAG 379

Qy    341 AACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGC 400

Db    380 CGATCGGAGCCGGAGACCAAGGTTTAATGTTCGGTTTTGCCTGCAACGAAACAAAGGAAC 439

Qy    401 TGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTA 460

Db    440 TTATGCCGCTGCCGATTTCGCTTGCCCATAAATTGTCTCGCCGTTTGACAGAAGTCCGCA 499

Qy    461 AAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATG 520

Db    500 AAGAGGAAATTCTGCCGTATCTGCGTCCGGACGGAAAGACACAGGTAACCGTTGAGTATG 559

Qy    521 ACGACGGCAAAATCG---TTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAG 577

Db    560 ATGAGAACAACAAGCCGATTCGCATCGATACAATCGTCATTTCGACTCAGCATCATCCTG 619

Qy    578 AGATCGACCAGAAATGGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGC 637

Db    620 AGATTTCGCTTGAACAAATCCAGCGCAATCTGAAAGAGCACGTCATCAATCCGGTAGTTC 679

Qy    638 CCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTA 697

Db    680 CAAAAGAGCTGATCGATGAAAACACGAAATACTTCATTAATCCTACGGGCCGTTTTGTTA 739

Qy    698 TCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACG 757

Db    740 TCGGCGGACCGCAGGGGGATGCCGGTTTAACAGGGCGCAAAATCATCGTTGACACCTATG 799

Qy    758 GCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACC 817

Db    800 GCGGCTATGCCCGTCACGGAGGCGGAGCGTTTTCCGGCAAGGATGCGACAAAGGTTGACA 859

Qy    818 GTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCG 877

Db    860 GATCTGCAGCTTATGCAGCGAGATATGTAGCGAAAAATATCGTCGGCGCAGGCCTTGCTG 919

Qy    878 ATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGG 937

Db    920 ATTCCTGTGAAGTACAGCTTGCCTACGCGATCGGTGTCGCTCAGCCGGTGTCCATTTCAA 979

Qy    938 TAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGT 997

Db    980 TCGATACATTCGGTACAGGCAAAGCGAGTGAAGAAACATTGATCGAAGTGGTTCGCAAAA 1039

Qy    998 TCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACA 1057

Db   1040 ACTTTGATCTAAGACCTGCCGGCATTATTAAAATGCTCGATCTTCGCCGTCCGATTTACA 1099

Qy   1058 AAGAAACCGCAGCATACGGTCACTTTGGTCG------TGAACATTTCCCGTGGGAAAAAA 1111

Db   1100 AACAAACGGCAGCTTACGGACATTTCGGCCGCCTCGATCTCGATCTGCCTTGGGAGCGCA 1159

Qy   1112 CCGACAAAGCGCAGCTGCTGCGCGATGCTGC 1142

Db   1160 CAGACAAAGCAGATCAGCTGAAGAAAGATGC 1190


RESULT 6
US-09-815-242-4516
; Sequence 4516, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
```

```
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4516
;   LENGTH: 1191
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-815-242-4516

  Query Match             32.5%;  Score 375.4;  DB 10;  Length 1191;
  Best Local Similarity   60.2%;  Pred. No. 7.2e-114;
  Matches  697;  Conservative    0;  Mismatches  421;  Indels   39;  Gaps    3;

Qy       7 AAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAA 66
Db      13 AAACGATTATTTACTTCAGAGTCTGTTACAGAAGGACACCCAGATAAAATCGCTGACCAA 72

Qy      67 ATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGC 126
Db      73 GTGTCAGATGCAATATTAGATGCTATTTTAAAAGACGACCCAAATGCACGTGTAGCTTGT 132

Qy     127 GAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGG 186
Db     133 GAAACAACTGTTACAACAGGTATGGCATTGATCGCCGGCGAAATTTCTACAACAACATAT 192

Qy     187 GTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGAC 246
Db     193 GTTGATATTCCGAAAGTTGTTAGAGAAACAATCAAAGAAATTGGATACACAAGAGCAAAA 252

Qy     247 ATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGAC 306
Db     253 TATGGTTATGATTATGAAACAATGGCAATTTTAACTGCGATTGATGAACAATCACCTGAC 312

Qy     307 ATCAACCAGGGCGTTGA----------------------------CCGTGCCGATCCG 336
Db     313 ATTGCACAAGGCGTGGATAAAGCATTAGAGTATCGTGATAAAGATAGCGAAGAAGAAATT 372

Qy     337 CTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGAC 396
Db     373 GAAGCAACAGGTGCAGGTGACCAAGGTTTAATGTTTGGTTACGCAACGAATGAAACAGAA 432

Qy     397 GTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTG 456
Db     433 ACGTATATGCCTCTAGCAATTTATTTATCACATCAATTGGCAAAACGTTTATCTGATGTG 492

Qy     457 CGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAG 516
Db     493 CGTAAAGATGGTACGTTAAATTATTTACGACCAGACGGTAAAGTTCAGGTCACTGTTGAA 552

Qy     517 TATGACG---ACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCT 573
Db     553 TATGATGAAAATGATAACCCAGTACGTATTGATACGATTGTCGTATCAACACAACATGCT 612

Qy     574 GAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATT 633
Db     613 GAGGATGTGACACTTGAACAAATTCAAGAAGACATCAAAGCTCATGTCATTTATCCAACT 672

Qy     634 CTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTC 693
Db     673 GTTCCAGAAAACTTGATTAATGAACAAACTAAATTCTACATCAATCCTACAGGACGCTTT 732

Qy     694 GTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACC 753
Db     733 GTAATCGGTGGACCTCAAGGTGATGCAGGATTAACAGGACGTAAAATCATTGTTGATACG 792

Qy     754 TACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTG 813
Db     793 TACGGTGGCTATGCACGTCATGGTGGCGGATGCTTCAGTGGTAAAGATCCTACAAAAGTA 852

Qy     814 GACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTG 873
Db     853 GACCGTTCAGCTGCATACGCTGCACGTTATGTGGCTAAAAATATTGTTGCAGCAGGTTTA 912

Qy     874 GCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATC 933
Db     913 GCAGATCAATGTGAAGTACAATTGGCATATGCGATTGGTGTTGCAGAACCAGTGTCAATT 972

Qy     934 ATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGT 993
Db     973 GCAATTGATACATTTGGAACAGGTAAAGTTTCTGAAGGACAACTTGTCGAAGCAGTTAGA 1032

Qy     994 GAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATC 1053
Db    1033 AAACACTTTGACCTAAGACCTGCAGGTATTATTAAAATGCTTGATTTGAAACAACCGATT 1092

Qy    1054 TACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACAT------TTCCCGTGGGAA 1107
Db    1093 TACAAACAAACTGCTGCTTATGGTCATTTTGGACGTACAGATGTCTTATTCCCATGGGAA 1152

Qy    1108 AAAACCGACAAAGCGCA 1124
Db    1153 AAATTAGACAAAGTGGA 1169


RESULT 7
US-09-815-242-8183
; Sequence 8183, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
```

```
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8183
;   LENGTH: 1296
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1296)
US-09-815-242-8183

  Query Match           32.5%;  Score 375.4;  DB 10;  Length 1296;
  Best Local Similarity  60.2%;  Pred. No. 7.5e-114;
  Matches  697;  Conservative    0;  Mismatches  421;  Indels   39;  Gaps    3;

Qy      7 AAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAA 66
          ||||   | ||||| || ||||| ||  | ||||| || || || ||||| |||||||||
Db    115 AAACGATTATTTACTTCAGAGTCTGTTACAGAAGGACACCCAGATAAAATCGCTGACCAA 174

Qy     67 ATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGC 126
           | || |||||  | ||||| || ||  |  || | || || || ||||| || |||||
Db    175 GTGTCAGATGCAATATTAGATGCTATTTTAAAAGACGACCCAAATGCACGTGTAGCTTGT 234

Qy    127 GAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGG 186
          |||||    || | ||| || |||| ||  | | |||||||||  | || |   | |
Db    235 GAAACAACTGTTACAACAGGTATGGCATTGATCGCCGGCGAAATTTCTACAACAACATAT 294

Qy    187 GTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGAC 246
          || || ||      |  |      |  | ||  |    |||||||| ||        |  |
Db    295 GTTGATATTCCGAAAGTTGTTAGAGAAACAATCAAAGAAATTGGATACACAAGAGCAAAA 354

Qy    247 ATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGAC 306
             || | |||   | |  |    ||  || | |  || || |   |||| || ||||||
Db    355 TATGGTTATGATTATGAAACAATGGCAATTTTAACTGCGATTGATGAACAATCACCTGAC 414

Qy    307 ATCAACCAGGGCGTTGA----------------------------CCGTGCCGATCCG 336
          ||    || ||||| ||                            |   |  ||
Db    415 ATTGCACAAGGCGTGGATAAAGCATTAGAGTATCGTGATAAAGATAGCGAAGAAGAAATT 474

Qy    337 CTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGAC 396
          | |   || || |||||||| ||| | |||||||| |||||||| |||||||| ||
Db    475 GAAGCAACAGGTGCAGGTGACCAAGGTTTAATGTTTGGTTACGCAACGAATGAAACAGAA 534

Qy    397 GTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTG 456
            . |||||   | | ||    |  |||| |   ||| | | |||    |||| |||
Db    535 ACGTATATGCCTCTAGCAATTTATTTATCACATCAATTGGCAAAACGTTTATCTGATGTG 594

Qy    457 CGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAG 516
          |||||| | || ||  |     |  | || || ||||  |||   ||||| ||| || |
Db    595 CGTAAAGATGGTACGTTAAATTATTTACGACCAGACGGTAAAGTTCAGGTCACTGTTGAA 654

Qy    517 TATGACG---ACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCT 573
          |||||  |   | |  ||    ||  |||| ||| |  | ||  | || || || ||  ||
Db    655 TATGATGAAAATGATAACCCAGTACGTATTGATACGATTGTCGTATCAACACAACATGCT 714

Qy    574 GAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATT 633
          || ||  |    |   ||    | |||||||   | |  |   |  ||||  | |||| |
Db    715 GAGGATGTGACACTTGAACAAATTCAAGAAGACATCAAAGCTCATGTCATTTATCCAACT 774

Qy    634 CTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTC 693
          | || |   | | | | | |      || ||||||| |||||| || || || || ||
Db    775 GTTCCAGAAAACTTGATTAATGAACAAACTAAATTCTACATCAATCCTACAGGACGCTTT 834

Qy    694 GTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACC 753
          || |||||||| ||    |||||    ||  | || || |||||||| || |||||||| 
Db    835 GTAATCGGTGGACCTCAAGGTGATGCAGGATTAACAGGACGTAAAATCATTGTTGATACG 894

Qy    754 TACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTG 813
          |||||  |||   || ||||| |||||||||    |||  |||||||||||||  |||||||
Db    895 TACGGTGGCTATGCACGTCATGGTGGCGGATGCTTCAGTGGTAAAGATCCTACAAAAGTA 954

Qy    814 GACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTG 873
          ||||||||| || || ||||| |||||||||||| || |||||| || ||||| || ||  |
Db    955 GACCGTTCAGCTGCATACGCTGCACGTTATGTGGCTAAAAATATTGTTGCAGCAGGTTTA 1014

Qy    874 GCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATC 933
          || ||||  |||||| | ||  |  | || || || || || || ||||    || ||
Db   1015 GCAGATCAATGTGAAGTACAATTGGCATATGCGATTGGTGTTGCAGAACCAGTGTCAATT 1074

Qy    934 ATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGT 993
            | || || || || || |  |||||  ||   | ||||||   |      || |
Db   1075 GCAATTGATACATTTGGAACAGGTAAAGTTTCTGAAGGACAACTTGTCGAAGCAGTTAGA 1134

Qy    994 GAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATC 1053
            |   ||| |||||  | ||    ||| | ||| | ||||| ||| ||   || |||||
Db   1135 AAACACTTTGACCTAAGACCTGCAGGTATTATTAAAATGCTTGATTTGAAACAACCGATT 1194

Qy   1054 TACAAAGAAACCGCAGCATACGGTCACTTCGGTCGTGAACAT------TTCCCGTGGGAA 1107
          |||||| |||| || || || ||||| ||||| ||| | ||      ||||| ||||||
Db   1195 TACAAACAAACTGCTGCTTATGGTCATTTTGGACGTACAGATGTCTTATTCCCATGGGAA 1254

Qy   1108 AAAACCGACAAAGCGCA 1124
          |||   ||||||| | |
Db   1255 AAATTAGACAAAGTGGA 1271


RESULT 8
US-09-815-242-8758
; Sequence 8758, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8758
;   LENGTH: 1296
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1296)
US-09-815-242-8758

  Query Match           32.5%;  Score 375.4;  DB 10;  Length 1296;
  Best Local Similarity  60.2%;  Pred. No. 7.5e-114;
```

```
  Matches  697;  Conservative    0;  Mismatches  421;  Indels   39;  Gaps    3;

Qy       7 AAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAA 66
           ||||   | ||||| || ||||| ||  | ||||| || || || ||||| |||||||||
Db     115 AAACGATTATTTACTTCAGAGTCTGTTACAGAAGGACACCCAGATAAAATCGCTGACCAA 174

Qy      67 ATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGC 126
             | || |||||  | ||||| || ||  |  || | ||||| || ||||| || |||||
Db     175 GTGTCAGATGCAATATTAGATGCTATTTTAAAAGACGATCCAAATGCACGTGTAGCTTGT 234

Qy     127 GAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGG 186
           |||||    || | ||| || ||||  ||  | | |||||||||  | || |   | |
Db     235 GAAACAACTGTTACAACAGGTATGGCATTGATCGCCGGCGAAATTTCTACAACAACATAT 294

Qy     187 GTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGAC 246
           || || ||      |  |      | || ||  ||   |||||||| ||         |  |
Db     295 GTTGATATTCCGAAAGTTGTTAGAGAAACAATTAAAGAAATTGGATACACAAGAGCAAAA 354

Qy     247 ATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGAC 306
             || | |||   | | |    ||  || | |  || || |   |||| || ||||||
Db     355 TATGGTTATGATTATGAAACAATGGCAATTTTAACTGCGATTGATGAACAATCACCTGAC 414

Qy     307 ATCAACCAGGGCGTTGA----------------------------CCGTGCCGATCCG 336
           ||    || |||||  ||                            |   |  ||
Db     415 ATTGCACAAGGCGTGGATAAAGCATTAGAGTATCGTGATAAAGATAGCGAAGAAGAAATT 474

Qy     337 CTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGAC 396
             | |   || || |||||||| ||| | |||||||| |||||||| ||||||||  ||
Db     475 GAAGCAACAGGTGCAGGTGACCAAGGTTTAATGTTTGGTTACGCAACGAATGAAACAGAG 534

Qy     397 GTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTG 456
             |   |||||   | | ||    |   |||| |   ||| | | |||    | || |||
Db     535 ACGTATATGCCTCTAGCAATTTATTTATCACATCAATTGGCAAAACGTTTATCAGATGTG 594

Qy     457 CGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAG 516
           |||||| | || ||  |·  |   | || || |||| |||   ||||| ||| || |
Db     595 CGTAAAGATGGTACGTTAAATTATTTACGACCAGACGGTAAAGTTCAGGTCACTGTTGAA 654

Qy     517 TATGACG---ACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCT 573
           ||||| |   | |  ||    ||  |||| ||| |  | ||  | || || || ||  ||
Db     655 TATGATGAAAATGATAACCCAGTACGTATTGATACGATTGTCGTATCAACACAACATGCT 714

Qy     574 GAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATT 633
           || ||  |    |  ||    | |||||||  | |  | | |  |||| | |||| |
Db     715 GAGGATGTGACACTTGAACAAATTCAAGAAGACATTAAAGCACATGTCATTTATCCAACT 774

Qy     634 CTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTC 693
           | || |   | | | | | |      || ||||||| |||||| || || || || ||
Db     775 GTTCCAGAAAACTTGATTAATGAACAAACTAAATTCTACATCAATCCTACAGGACGCTTT 834

Qy     694 GTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACC 753
           || |||||||| ||    |||||    ||  | || || |||||||| || ||||||||
Db     835 GTAATCGGTGGACCTCAAGGTGATGCAGGATTAACAGGACGTAAAATCATTGTTGATACG 894

Qy     754 TACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTG 813
           ||||| |||   || ||||| ||||||||    |||  |||||| |||||  |||||||
Db     895 TACGGTGGCTATGCACGTCATGGTGGCGGATGCTTCAGTGGTAAGGATCCTACAAAAGTA 954

Qy     814 GACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTG 873
           |||||||| || || ||||| |||||||||| || ||||| || ||||| || ||  |
Db     955 GACCGTTCAGCTGCATACGCTGCACGTTATGTGGCTAAAAATATTGTTGCAGCAGGTTTA 1014

Qy     874 GCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATC 933
           || ||||  |||||| | ||  |  | || || || || || || ||||    || ||
Db    1015 GCAGATCAATGTGAAGTACAATTGGCATATGCGATTGGTGTTGCAGAACCAGTGTCAATT 1074

Qy     934 ATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGT 993
             | || || || || || | |||||  ||   | |||||   |       ||  |
Db    1075 GCAATTGATACATTTGGAACAGGTAAAGTTTCTGAAGGGCAACTTGTCGAAGCAGTTAGA 1134

Qy     994 GAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATC 1053
             |   ||| |||||  | ||    ||| | ||| | ||||| ||| ||   || |||||
Db    1135 AAACACTTTGACCTAAGACCTGCAGGTATTATTAAAATGCTTGATTTGAAACAACCGATT 1194

Qy    1054 TACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACAT------TTCCCGTGGGAA 1107
           |||||| |||| || || || ||||| ||||| |||  | ||      ||||| ||||||
Db    1195 TACAAACAAACTGCTGCTTATGGTCATTTTGGACGTACAGATGTCTTATTCCCATGGGAA 1254

Qy    1108 AAAACCGACAAAGCGCA 1124
           |||   |||||||  | |
Db    1255 AAATTAGACAAAGTGGA 1271


RESULT 9
US-08-781-986A-143/c
; Sequence 143, Application US/08781986A
; Publication No. US20030054436A1
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,936A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Bob
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PB248PP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  143:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10320 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-781-986A-143

  Query Match             32.5%;  Score 375.4;  DB 7;  Length 10320;
  Best Local Similarity   60.2%;  Pred. No. 2.1e-113;
  Matches  697;  Conservative    0;  Mismatches  421;  Indels   39;  Gaps    3;

Qy       7 AAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAA 66
           ||||   | ||||| || ||||| ||  | ||||| || || || ||||| |||||||||
Db    2744 AAACGATTATTTACTTCAGAGTCTGTTACAGAAGGACACCCAGATAAAATCGCTGACCAA 2685

Qy      67 ATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGC 126
             | || |||||  | ||||| || ||  |  || | ||||| || ||||| || |||||
Db    2684 GTGTCAGATGCAATATTAGATGCTATTTTAAAAGACGATCCAAATGCACGTGTAGCTTGT 2625

Qy     127 GAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGG 186
           |||||    || | ||| || ||||  ||  | | |||||||||  | || |   | |
Db    2624 GAAACAACTGTTACAACAGGTATGGCATTGATCGCCGGCGAAATTTCTACAACAACATAT 2565

Qy     187 GTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGAC 246
           || || ||      |  |      | || ||  ||   |||||||| ||         |  |
Db    2564 GTTGATATTCCGAAAGTTGTTAGAGAAACAATTAAAGAAATTGGATACACAAGAGCAAAA 2505

Qy     247 ATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGAC 306
```

```
Db   2504 TATGGTTATGATTATGAAACAATGGCAATTTTAACTGCGATTGATGAACAATCACCTGAC 2445

Qy    307 ATCAACCAGGGCGTTGA----------------------------CCGTGCCGATCCG 336
Db   2444 ATTGCACAAGGCGTGGATAAAGCATTAGAGTATCGTGATAAAGATAGCGAAGAAGAAATT 2385

Qy    337 CTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGAC 396
Db   2384 GAAGCAACAGGTGCAGGTGACCAAGGTTTAATGTTTGGTTACGCAACGAATGAAACAGAG 2325

Qy    397 GTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAACTG 456
Db   2324 ACGTATATGCCTCTAGCAATTTATTTATCACATCAATTGGCAAAACGTTTATCAGATGTG 2265

Qy    457 CGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAG 516
Db   2264 CGTAAAGATGGTACGTTAAATTATTTACGACCAGACGGTAAAGTTCAGGTCACTGTTGAA 2205

Qy    517 TATGACG---ACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCT 573
Db   2204 TATGATGAAAATGATAACCCAGTACGTATTGATACGATTGTCGTATCAACACAACATGCT 2145

Qy    574 GAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATT 633
Db   2144 GAGGATGTGACACTTGAACAAATTCAAGAAGACATTAAAGCACATGTCATTTATCCAACT 2085

Qy    634 CTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTC 693
Db   2084 GTTCCAGAAAACTTGATTAATGAACAAACTAAATTCTACATCAATCCTACAGGACGCTTT 2025

Qy    694 GTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACC 753
Db   2024 GTAATCGGTGGACCTCAAGGTGATGCAGGATTAACAGGACGTAAAATCATTGTTGATACG 1965

Qy    754 TACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTG 813
Db   1964 TACGGTGGCTATGCACGTCATGGTGGCGGATGCTTCAGTGGTAAGGATCCTACAAAAGTA 1905

Qy    814 GACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTG 873
Db   1904 GACCGTTCAGCTGCATACGCTGCACGTTATGTGGCTAAAAATATTGTTGCAGCAGGTTTA 1845

Qy    874 GCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATC 933
Db   1844 GCAGATCAATGTGAAGTACAATTGGCATATGCGATTGGTGTTGCAGAACCAGTGTCAATT 1785

Qy    934 ATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGT 993
Db   1784 GCAATTGATACATTTGGAACAGGTAAAGTTTCTGAAGGGCAACTTGTCGAAGCAGTTAGA 1725

Qy    994 GAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATC 1053
Db   1724 AAACACTTTGACCTAAGACCTGCAGGTATTATTAAAATGCTTGATTTGAAACAACCGATT 1665

Qy   1054 TACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACAT------TTCCCGTGGGAA 1107
Db   1664 TACAAACAAACTGCTGCTTATGGTCATTTTGGACGTACAGATGTCTTATTCCCATGGGAA 1605

Qy   1108 AAAACCGACAAAGCGCA 1124
Db   1604 AAATTAGACAAAGTGGA 1588


RESULT 10
US-09-943-702-1
; Sequence 1, Application US/09943702
; Publication No. US20030022322A1
;   GENERAL INFORMATION:
;        APPLICANT: DeHoff, Bradley S.
;                   Rosteck Jr., Paul R.
;        TITLE OF INVENTION: SAM Operon
;        NUMBER OF SEQUENCES: 6
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Eli Lilly and Company
;             STREET: Lilly Corporate Center
;             CITY: Indianapolis
;             STATE: Indiana
;             COUNTRY: U.S.
;             ZIP: 46285
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/943,702
;             FILING DATE: 31-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/955,957
;             FILING DATE: 1997-10-22
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Webster, Thomas D.
;             REGISTRATION NUMBER: 39,072
;             REFERENCE/DOCKET NUMBER: P-10162
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 317-276-3334
;   INFORMATION FOR SEQ ID NO: 1:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4848 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  986..2209
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  2241..3341
;        SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-943-702-1

  Query Match           32.4%;  Score 374;  DB 9;  Length 4848;
  Best Local Similarity  59.7%;  Pred. No. 4.2e-113;
  Matches  728;  Conservative    0;  Mismatches  420;  Indels   72;  Gaps    3;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
Db     986 ATGTCACGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCCACCCCGACAAGATCGCC 1045

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCCATCCTCGAACAGGATCCGAAAGCACGCGTT 120
Db    1046 GACCGGATCAGCGACACCGTCCTCGATGCCCTGCTCGCCCGGGACCCGCGGGCCAGGGTC 1105

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
Db    1106 GCCGTCGAGACCCTGATCACCACCGGTCAGGTCCATATCGCCGGCGAGGTCACCACCACC 1165

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
Db    1166 GCGTACGCGCCCATCGCCCAACTGGTGCGCGACACCGTCCTGTCCATCGGCTACGACTCC 1225

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
Db    1226 TCGGCCAAGGGCTTCGACGGCGCCTCGTGCGGGGTGTCGGTCTCCATCGGCGCGCAGTCC 1285

Qy     301 CCTGACATCAACCAGGGCGTTGAC---------------------------------------- 324
Db    1286 CCGGACATCGCCCGGGGTGTGGACACCGCGTACGAGCGGCGGGGCGGGGGCACGGCCCCG 1345

Qy     325 ---------CGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGC 375
```

```
Db   1346 GGCGGACCGGGTGACGAGCTGGACCGGCAGGGCGCGGGCGACCAGGGCCTGATGTTCGGC 1405

Qy    376 TACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTG 435

Db   1406 TACGCCTGCGACGAGACCCCCGAGCTGATGCCGCTGCCGATCAACCTGGCGCACCGGCTC 1465

Qy    436 GTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCG 495

Db   1466 TCCCGGCGGCTGTCGGAGGTGCGGAAGAACGGCACGATCCCCTACCTCCGCCCCGACGGC 1525

Qy    496 AAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTG 555

Db   1526 AAGACCCAGGTCACCATCGAGTACGACGGCGACAAGGCGGTCCGCCTCGACACGGTCGTG 1585

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615

Db   1586 GTCTCCTCCCAGCACGCCTCCGGCATCGACCTGGACTCGCTACTGGCGCCCGACATCCGC 1645

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCA-- 673

Db   1646 CGGCATGTCGTGGAGCCCGTCCTCGCCGGACTGGCCGAGGACGGCATCAAGCTGGACACC 1705

Qy    674 ------------------TCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714

Db   1706 GCGGGCTACCGGCTGCTGGTCAATCCCACCGGCCGGTTCGAGATCGGCGGCCCGATGGGC 1765

Qy    715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774

Db   1766 GACGCGGGCCTGACCGGCCGCAAGATCATCATCGACACCTACGGCGGCATGGCCCGGCAT 1825

Qy    775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834

Db   1826 GGCGGCGGTGCCTTCTCCGGCAAGGACCCGTCCAAGGTGGACCGTTCGGCGGCGTACGCA 1885

Qy    835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894

Db   1886 ATGCGCTGGGTGGCCAAGAACGTCGTGGCGGCGGGGCTGGCCTCGCGGTGTGAGGTGCAG 1945

Qy    895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 954

Db   1946 GTCGCCTACGCCATCGGCAAGGCCGAGCCGGTCGGTCTGTTCGTGGAGACCTTCGGCACC 2005

Qy    955 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCA 1014

Db   2006 GCCACCGTGGACGTGGAGCGGATCGAGCAGGCGATCGGCGAGGTCTTCGACCTCCGCCCG 2065

Qy   1015 TACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1074

Db   2066 GCCGCGATCATCCGGGATCTGGACCTGCTCCGCCCGATCTACGCCAAGACCGCCGCCTAC 2125

Qy   1075 GGTCACTTTGGTCGTGAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTG 1128

Db   2126 GGCCACTTCGGCCGCGAACTGCCCGAGTTCACCTGGGAACGCACCGACCGCACCGAGCAG 2185

Qy   1129 CTGCGCGATGCTGCCGGTCT 1148

Db   2186 CTCATCGCCGCGGCCGGCCT 2205


RESULT 11
US-09-943-702-4
; Sequence 4, Application US/09943702
; Publication No. US20030022322A1
;   GENERAL INFORMATION:
;        APPLICANT: DeHoff, Bradley S.
;                   Rosteck Jr., Paul R.
;        TITLE OF INVENTION: SAM Operon
;        NUMBER OF SEQUENCES: 6
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Eli Lilly and Company
;             STREET: Lilly Corporate Center
;             CITY: Indianapolis
;             STATE: Indiana
;             COUNTRY: U.S.
;             ZIP: 46285
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/943,702
;             FILING DATE: 31-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/955,957
;             FILING DATE: 1997-10-22
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Webster, Thomas D.
;             REGISTRATION NUMBER: 39,872
;             REFERENCE/DOCKET NUMBER: P-10162
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 317-276-3334
;   INFORMATION FOR SEQ ID NO: 4:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4848 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: DNA (genomic)
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        FEATURE:
;             NAME/KEY:  CDS
;             LOCATION:  3338..4255
;        SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-09-943-702-4

  Query Match             32.4%;  Score 374;  DB 9;  Length 4848;
  Best Local Similarity   59.7%;  Pred. No. 4.2e-113;
  Matches  728;  Conservative    0;  Mismatches  420;  Indels   72;  Gaps    3;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Db    986 ATGTCACGTCGCCTGTTCACCTCGGAGTCCGTGACCGAGGGCCACCCCGACAAGATCGCC 1045

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Db   1046 GACCGGATCAGCGACACCGTCCTCGATGCCCTGCTCGCCCGGGACCCGCGGGCCAGGGTC 1105

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Db   1106 GCCGTCGAGACCCTGATCACCACCGGTCAGGTCCATATCGCCGGCGAGGTCACCACCACC 1165

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Db   1166 GCGTACGCGCCCATCGCCCAACTGGTGCGCGACACCGTCCTGTCCATCGGCTACGACTCC 1225

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Db   1226 TCGGCCAAGGGCTTCGACGGCGCCTCGTGCGGGGTGTCGGTCTCCATCGGCGCGCAGTCC 1285

Qy    301 CCTGACATCAACCAGGGCGTTGAC---------------------------------------- 324

Db   1286 CCGGACATCGCCCGGGGTGTGGACACCGCGTACGAGCGGCGGGGCGGGGGCACGGCCCCG 1345

Qy    325 ---------CGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGC 375

Db   1346 GGCGGACCGGGTGACGAGCTGGACCGGCAGGGCGCGGGCGACCAGGGCCTGATGTTCGGC 1405

Qy    376 TACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTG 435

Db   1406 TACGCCTGCGACGAGACCCCCGAGCTGATGCCGCTGCCGATCAACCTGGCGCACCGGCTC 1465

Qy    436 GTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCG 495
```

```
Db   1466 TCCCGGCCGCTGTCGGAGGTGCGGAAGAACGGCACGATCCCCTACCTCCGCCCCGACGGC 1525

Qy    496 AAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTG 555

Db   1526 AAGACCCAGGTCACCATCGAGTACGACGGCGACAAGGCGGTCCGCCTCGACACGGTCGTG 1585

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615

Db   1586 GTCTCCTCCCAGCACGCCTCCGGCATCGACCTGGACTCGCTACTGGCGCCCGACATCCGC 1645

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCA-- 673

Db   1646 CGGCATGTCGTGGAGCCCGTCCTCGCCGGACTGGCCGAGGACGGCATCAAGCTGGACACC 1705

Qy    674 ------------------TCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714

Db   1706 GCGGGCTACCGGCTGCTGGTCAATCCCACCGGCCGGTTCGAGATCGGCGGCCCGATGGGC 1765

Qy    715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774

Db   1766 GACGCGGGCCTGACCGGCCGCAAGATCATCATCGACACCTACGGCGGCATGGCCCGGCAT 1825

Qy    775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834

Db   1826 GGCGGCGGTGCCTTCTCCGGCAAGGACCCGTCCAAGGTGGACCGTTCGGCGGCGTACGCA 1885

Qy    835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894

Db   1886 ATGCGCTGGGTGGCCAAGAACGTCGTGGCGGCGGGGCTGGCCTCGCGGTGTGAGGTGCAG 1945

Qy    895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 954

Db   1946 GTCGCCTACGCCATCGGCAAGGCCGAGCCGGTCGGTCTGTTCGTGGAGACCTTCGGCACC 2005

Qy    955 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCA 1014

Db   2006 GCCACCGTGGACGTGGAGCGGATCGAGCAGGCGATCGGCGAGGTCTTCGACCTCCGCCCG 2065

Qy   1015 TACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1074

Db   2066 GCCGCGATCATCCGGGATCTGGACCTGCTCCGCCCGATCTACGCCAAGACCGCCGCCTAC 2125

Qy   1075 GGTCACTTTGGTCGTGAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTG 1128

Db   2126 GGCCACTTCGGCCGCGAACTGCCCGAGTTCACCTGGGAACGCACCGACCGCACCGAGCAG 2185

Qy   1129 CTGCGCGATGCTGCCGGTCT 1148

Db   2186 CTCATCGCCGCGGCCGGCCT 2205


RESULT 12
US-09-943-702-6
; Sequence 6, Application US/09943702
; Publication No. US20030022322A1
;   GENERAL INFORMATION:
;        APPLICANT: DeHoff, Bradley S.
;                   Rosteck Jr., Paul R.
;        TITLE OF INVENTION: SAM Operon
;        NUMBER OF SEQUENCES: 6
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Eli Lilly and Company
;             STREET: Lilly Corporate Center
;             CITY: Indianapolis
;             STATE: Indiana
;             COUNTRY: U.S.
;             ZIP: 46285
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Floppy disk
;             COMPUTER: IBM PC compatible
;             OPERATING SYSTEM: PC-DOS/MS-DOS
;             SOFTWARE: PatentIn Release #1.0, Version #1.30
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/943,702
;             FILING DATE: 31-Aug-2001
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 08/955,957
;             FILING DATE: 1997-10-22
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Webster, Thomas D.
;             REGISTRATION NUMBER: 39,872
;             REFERENCE/DOCKET NUMBER: P-10162
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: 317-276-3334
;   INFORMATION FOR SEQ ID NO: 6:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 4848 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: single
;             TOPOLOGY: linear
;        MOLECULE TYPE: mRNA
;        HYPOTHETICAL: NO
;        ANTI-SENSE: NO
;        SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-943-702-6

  Query Match             32.4%;  Score 374;  DB 9;  Length 4848;
  Best Local Similarity   49.7%;  Pred. No. 4.2e-113;
  Matches  606;  Conservative  122;  Mismatches  420;  Indels   72;  Gaps    3;

Qy      1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60

Db    986 AUGUCACGUCGCCUGUUCACCUCGGAGUCCGUGACCGAGGGCCACCCCGACAAGAUCGCC 1045

Qy     61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120

Db   1046 GACCGGAUCAGCGACACCGUCCUCGAUGCCCUGCUCGCCCGGGACCCGCGGGCCAGGGUC 1105

Qy    121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180

Db   1106 GCCGUCGAGACCCUGAUCACCACCGGUCAGGUCCAUAUCGCCGGCGAGGUCACCACCACC 1165

Qy    181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240

Db   1166 GCGUACGCGCCCAUCGCCCAACUGGUGCGCGACACCGUCCUGUCCAUCGGCUACGACUCC 1225

Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300

Db   1226 UCGGCCAAGGGCUUCGACGGCGCCUCGUGCGGGGUGUCGGUCUCCAUCGGCGCGCAGUCC 1285

Qy    301 CCTGACATCAACCAGGGCGTTGAC---------------------------------------- 324

Db   1286 CCGGACAUCGCCCGGGGUGUGGACACCGCGUACGAGCGGCGGGGCGGGGGCACGGCCCCG 1345

Qy    325 ---------CGTGCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGC 375

Db   1346 GGCGGACCGGGUGACGAGCUGGACCGGCAGGGCGCGGGCGACCAGGGCCUGAUGUUCGGC 1405

Qy    376 TACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTG 435

Db   1406 UACGCCUGCGACGAGACCCCCGAGCUGAUGCCGCUGCCGAUCAACCUGGCGCACCGGCUC 1465

Qy    436 GTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCG 495

Db   1466 UCCCGGCGGCUGUCGGAGGUGCGGAAGAACGGCACGAUCCCCUACCUCCGCCCCGACGGC 1525

Qy    496 AAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTG 555

Db   1526 AAGACCCAGGUCACCAUCGAGUACGACGGCGACAAGGCGGUCCGCCUCGACACGGUCGUG 1585

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615

Db   1586 GUCUCCUCCCAGCACGCCUCCGGCAUCGACCUGGACUCGCUACUGGCGCCCGACAUCCGC 1645

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCA-- 673
```

```
Db   1646 CGGCAUGUCGUGGAGCCCGUCCUCGCCGGACUGGCCGAGGACGGCAUCAAGCUGGACACC 1705

Qy    674 ------------------TCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714
                            :||| || ||||| || ::||  |:||| ||||| |:|||
Db   1706 GCGGGCUACCGGCUGCUGGUCAAUCCCACCGGCCGGUUCGAGAUCGGCGGCCCGAUGGGC 1765

Qy    715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774
          |||   || |:||| || || || |: |:| : || |||:|||||||||:||| || ||
Db   1766 GACGCGGGCCUGACCGGCCGCAAGAUCAUCAUCGACACCUACGGCGGCAUGGCCCGGCAU 1825

Qy    775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834
          || |||||:|| ::|:| || || || || :| || |:||||||::| || || :|||||
Db   1826 GGCGGCGGUGCCUUCUCCGGCAAGGACCCGUCCAAGGUGGACCGUUCGGCGGCGUACGCA 1885

Qy    835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894
            ||  :  |: || || ||| :||: || || || |:|||   || :|:|| : |||
Db   1886 AUGCGCUGGGUGGCCAAGAACGUCGUGGCGGCGGGGCUGGCCUCGCGGUGUGAGGUGCAG 1945

Qy    895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACT 954
          |:  ||:|||| |:|||  ||| || |||  |    :  : |: || || ::||| ||
Db   1946 GUCGCCUACGCCAUCGGCAAGGCCGAGCCGGUCGGUCUGUUCGUGGAGACCUUCGGCACC 2005

Qy    955 GAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCA 1014
          |  |  |:|     || |   :    | |  | :  | ||| :|::|||||: |||||
Db   2006 GCCACCGUGGACGUGGAGCGGAUCGAGCAGGCGAUCGGCGAGGUCUUCGACCUCCGCCCG 2065

Qy   1015 TACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATAC 1074
            ||   : |: | |   |:||| |:||: | ||||||:|:||    | ||||| || :||
Db   2066 GCCGCGAUCAUCCGGGAUCUGGACCUGCUCCGCCCGAUCUACGCCAAGACCGCCGCCUAC 2125

Qy   1075 GGTCACTTTGGTCGTGAAC------ATTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTG 1128
          || |||:: || || ||||      | ::| | :|||||   ||||||    |  ||| |
Db   2126 GGCCACUUCGGCCGCGAACUGCCCGAGUUCACCUGGGAACGCACCGACCGCACCGAGCAG 2185

Qy   1129 CTGCGCGATGCTGCCGGTCT 1148
          |:   ||  || ||||| |:
Db   2186 CUCAUCGCCGCGGCCGGCCU 2205


RESULT 13
US-09-815-242-6669
; Sequence 6669, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6669
;   LENGTH: 1185
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1185)
US-09-815-242-6669

  Query Match             32.1%;  Score 370.6;  DB 10;  Length 1185;
  Best Local Similarity   60.0%;  Pred. No. 2.8e-112;
  Matches  688;  Conservative    0;  Mismatches  429;  Indels   30;  Gaps    3;

Qy       4 GCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGAC 63
           | || |||  | ||||| || || ||||| || ||||| ||||| || ||  | ||||||
Db       7 GAAAGACATTTATTTACATCAGAATCCGTTTCAGAAGGACATCCAGATAAGGTAGCTGAC 66

Qy      64 CAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCT 123
           ||| |   ||||||  ||||||| ||||| | ||  | |||||||  || ||||| ||
Db      67 CAAGTCAGTGATGCAATTTTAGATGCGATTTTGGAGAAAGATCCGATGGCCCGCGTGGCA 126

Qy     124 TGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCC 183
           || ||||| | ||||| ||| ||  | ||  |||| || ||||||  | || |  ||
Db     127 TGTGAAACATCCGTAACAACTGGTTTAGTCCTAGTATTTGGAGAAATTTCTACAACAGCT 186

Qy     184 TGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCC 243
           |  ||||| ||| || | ||     ||| | || ||     || ||||| |||   |  |
Db     187 TATGTAGATATCCAAAAAATTGTTCGTCAAACTGTGAAAGATATTGGTTATACACGAGCA 246

Qy     244 GACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCT 303
           |    || ||||| |   |  |    || |||||    || ||||   | || || |||
Db     247 AAATATGGATTTGATGGAGAGACTGTCGCAGTTCTTGTAGCAATCGATGAGCAATCACCT 306

Qy     304 GACATCAACCAGGGCGTTGACCGTGCC--------------------GATCCGCTGGAA 342
           || |||   || || ||||||  |||                    ||     ||
Db     307 GATATCGCTCAAGGGGTTGACGCAGCCTTAGAAGTGCGTGATCAAGATGAAAAAGACGAT 366

Qy     343 CAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTG 402
             ||||| |||||||| ||  |  |||||||  |  ||   | || |||||    |  |
Db     367 ATTGGCGCTGGTGACCAAGGCTTACTGTTTGGTTTTGCCGTTGACGAAACACCAGAATTA 426

Qy     403 ATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAA 462
           ||||||  ||| ||| | |     || ||| |||| |  |||  || ||| |||| ||
Db     427 ATGCCATTACCGATCGCGTTAAGTCATCGTTTGGTTCGCCGTTTAGCAGAATTGCGGAAG 486

Qy     463 AACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGAC 522
           |     | || |   | || || || || |||   || || ||  |  | ||||
Db     487 GAAAAAGTCTTACCTTACTTACGTCCTGATGCAAAATCTCAAGTAACAGTGGAATATGAT 546

Qy     523 GAC---GGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAG 579
           ||    ||  ||       || | ||| |  | ||  |    || || ||   ||| ||
Db     547 GATCAAGGTCAACCACAACGTGTGGATACGATTGTCATCAGTACACAACATGATGATGAA 606

Qy     580 ATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCC 639
           |    |    ||    |  || |||   | |  ||| |  | |||||     |  | ||
Db     607 ACAACCTTAGAACAAATCGAAAAAGACATTAAAGAACAAGTAATCAATGAGGTGATTCCT 666

Qy     640 GCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATC 699
             ||||   |  |  ||  || |||| ||| |||||||| || ||||| || || |||
Db     667 CATGAACTTTTAGATGATGAAACAAAATACTTTATCAACCCAACTGGTCGCTTTGTAATC 726

Qy     700 GGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGC 759
           ||||||||    || ||   |||  | ||||||||||||||||||||||||||| || |||
Db     727 GGTGGCCCTCAAGGAGATGCCGGCTTAACTGGTCGTAAAATTATCGTTGATACTTATGGC 786

Qy     760 GGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGT 819
           ||    || ||||| || |||||||| || ||||| |||||| || | ||||| |||||
Db     787 GGTTATGCTCGTCATGGCGGCGGTGCTTTTTCTGGAAAAGATGCAACCAAAGTTGACCGC 846

Qy     820 TCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGAT 879
           || ||    || || || |||||| |||||||||||| ||||||||||  ||||  |
Db     847 TCTGCTAGTTATGCTGCGCGTTATATCGCGAAAAACATTGTTGCTGCTGGTTTGGCGAAA 906
```

```
Qy    880 CGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTA 939
Db    907 AAAGTAGAAGTCCAATTAGCTTATGCAATTGGCGTGGCGCAACCAGTGTCTATTTCTATT 966

Qy    940 GAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTC 999
Db    967 AATACGTTTGGTACGAGTGCTCTTCCAGAAAGTAAATTAATCGAAGCAGTTCGTAAAAAC 1026

Qy   1000 TTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAA 1059
Db   1027 TTTGATTTACGTCCAGCAGGTATTATTGAAATGCTAGATTTACGTCGCCCAATTTATAAA 1086

Qy   1060 GAAACCGCAGCATACGGTCACTTTGGTCG------TGAACATTTCCCGTGGGAAAAAACC 1113
Db   1087 CAAACTGCGGCTTATGGTCATTTTGGTCGAACAGATATTGATTTACCGTGGGAACAAACA 1146

Qy   1114 GACAAAG 1120
Db   1147 GATAAAG 1153


RESULT 14
US-09-815-242-9315
; Sequence 9315, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9315
;  LENGTH: 1191
;  TYPE: DNA
;  ORGANISM: Streptococcus pneumoniae
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(1191)
US-09-815-242-9315

  Query Match             32.0%;  Score 369.8;  DB 10;  Length 1191;
  Best Local Similarity   60.6%;  Pred. No. 5.1e-112;
  Matches  677;  Conservative    0;  Mismatches  407;  Indels   33;  Gaps    3;

Qy     14 TTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTCTG 73
Db     17 TATTCACGTCTGAATCTGTATCTGAGGGGCATCCGGATAAGATTGCAGACCAAATTTCAG 76

Qy     74 ATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAACCT 133
Db     77 ATGCGATTTTGGATGCTATTTTAGCAAAGGATCCAGAGGCGCACGTTGCTGCTGAAACAG 136

Qy    134 ACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGACA 193
Db    137 CTGTATATACTGGTTCTGTCCACGTTTTTGGTGAAATTTCTACAAATGCCTATGTGGATA 196

Qy    194 TCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATGGGCT 253
Db    197 TTAACCGTGTGGTTCGTGATACCATTGCAGAGATTGGTTATACCAATACAGAATATGGAT 256

Qy    254 TTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAACC 313
Db    257 TTTCTGCTGAGACGGTGGGAGTACACCCATCTTTGGTGGAACAATCTCCTGACATCGCTC 316

Qy    314 AGGGCGTTGACCGTGCC----------------------------GATCCGCTGGAACAG- 345
Db    317 AAGGTGTTAACGAAGCCTTGGAGGTTCGCGGAAATGCTGATCAAGATCCACTGGACTTGA 376

Qy    346 --GGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGA 403
Db    377 TTGGAGCAGGTGACCAAGGGCTCATGTTTGGATTTGCAGTAGATGAAACAGAAGAGCTTA 436

Qy    404 TGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAA 463
Db    437 TGCCATTGCCAATTGCACTCAGTCATAAATTGGTTCGTCGTCTGGCAGAACTTCGTAAGT 496

Qy    464 ACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACG 523
Db    497 CTGGAGAAATTAGCTATCTCCGTCCAGATGCAAAATCACAAGTTACAGTTGAGTACGATG 556

Qy    524 AC---GGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGA 580
Db    557 AAAATGACCGTCCGGTACGTGTAGATACAGTCGTTATTTCTACTCAGCATGATCCAGAGG 616

Qy    581 TCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCG 640
Db    617 CCACTAATGAACAAATCCATCAAGATGTGATTGACAAGGTCATCAAAGAAGTTATTCCAT 676

Qy    641 CTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCG 700
Db    677 CTTCTTATCTTGATGATAAGACAAAATTCTTTATCAATCCGACAGGTCGTTTTGTAATCG 736

Qy    701 GTGGCCCAATGGGTGACTGCGGTCTGACTCGTCGTAAAATTATCGTTGATACCTACGGCG 760
Db    737 GTGGTCCTCAAGGGGACTCAGGTTTGACTCGTCGTAAGATTATTGTAGATACTTATGGTG 796

Qy    761 GCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTT 820
Db    797 GCTACTCTCGTCATGGTGGTGGTGCCTTCTCTGGTAAAGATGCGACTAAGGTGGATCGTT 856

Qy    821 CCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATC 880
Db    857 CAGCCTCTTATGCGGCTCGCTATATTGCCAAGAATATCGTTGCAGCAGGCCTTGCTAAGA 916

Qy    881 GTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAG 940
Db    917 AGGCAGAAGTGCAGTTGGCCTATGCTATCGGTGTTGCGCAACCTGTTTCTGTTCGTATCG 976

Qy    941 AAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCT 1000
Db    977 ATACTTTCGGTACAGGAACAGTAGCTGAAAGTCAACTTGAAAAAGCGGCTCGTCAAATCT 1036

Qy   1001 TCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAG 1060
Db   1037 TTGACCTTCGCCCTGCAGGGATTATCCAAATGCTGGACCTCAAGCGTCCAATTTACCGTC 1096

Qy   1061 AAACCGCAGCATACGGTCACTTTGGTCGTGAACATTT 1097
Db   1097 AAACATCGGCTTACGGTCACATGGGACGTACAGATAT 1133
```

```
RESULT 15
US-09-815-242-3823
; Sequence 3823, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
;  APPLICANT: Haselbeck, Robert
;  APPLICANT:  Ohlsen, Kari L.
;  APPLICANT:  Zyskind, Judith W.
;  APPLICANT:  Wall, Daniel
;  APPLICANT:  Trawick, John D.
;  APPLICANT:  Carr, Grant J.
;  APPLICANT:  Yamamoto, Robert T.
;  APPLICANT:  Xu, H. Howard
;  TITLE OF INVENTION: Identification of Essential Genes in
;  TITLE OF INVENTION:  Prokaryotes
;  FILE REFERENCE: ELITRA.011A
;  CURRENT APPLICATION NUMBER: US/09/815,242
;  CURRENT FILING DATE:  2001-03-21
;  PRIOR APPLICATION NUMBER: 60/191,078
;  PRIOR FILING DATE: 2000-03-21
;  PRIOR APPLICATION NUMBER: 60/206,848
;  PRIOR FILING DATE: 2000-05-23
;  PRIOR APPLICATION NUMBER: 60/207,727
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: 60/242,578
;  PRIOR FILING DATE: 2000-10-23
;  PRIOR APPLICATION NUMBER: 60/253,625
;  PRIOR FILING DATE: 2000-11-27
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3823
;   LENGTH: 1158
;   TYPE: DNA
;   ORGANISM: Enterococcus faecalis
US-09-815-242-3823

  Query Match             32.0%;  Score 369.2;  DB 10;  Length 1158;
  Best Local Similarity   60.0%;  Pred. No. 8e-112;
  Matches  686;  Conservative    0;  Mismatches  428;  Indels   30;  Gaps    3;

Qy       7 AAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAA 66
Db       1 AGACATTTATTTACATCAGAATCCGTTTCAGAAGGACATCCAGATAAGGTAGCTGACCAA 60

Qy      67 ATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGC 126
Db      61 GTCAGTGATGCAATTTTAGATGCGATTTTGGAGAAAGATCCGATGGCCCGCGTGGCATGT 120

Qy     127 GAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGG 186
Db     121 GAAACATCCGTAACAACTGGTTTAGTCCTAGTATTTGGAGAAATTTCTACAACAGCTTAT 180

Qy     187 GTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGAC 246
Db     181 GTAGATATCCAAAAAATTGTTCGTCAAACTGTGAAAGATATTGGTTATACACGAGCAAAA 240

Qy     247 ATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGAC 306
Db     241 TATGGATTTGATGGAGAGACTGTCGCAGTTCTTGTAGCAATCGATGAGCAATCACCTGAT 300

Qy     307 ATCAACCAGGGCGTTGACCGTGCC----------------------GATCCGCTGGAACAG 345
Db     301 ATCGCTCAAGGGGTTGACGCAGCCTTAGAAGTGCGTGATCAAGATGAAAAAGACGATATT 360

Qy     346 GGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATG 405
Db     361 GGCGCTGGTGACCAAGGCTTACTGTTTGGTTTTGCCGTTGACGAAACACCAGAATTAATG 420

Qy     406 CCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAAC 465
Db     421 CCATTACCGATCGCGTTAAGTCATCGTTTGGTTCGCCGTTTAGCAGAATTGCGGAAGGAA 480

Qy     466 GGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACG-- 523
Db     481 AAAGTCTTACCTTACTTACGTCCTGATGCAAAATCTCAAGTAACAGTGGAATATGATGAT 540

Qy     524 -ACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATC 582
Db     541 CAAGGTCAACCACAACGTGTGGATACGATTGTCATCAGTACACAACATGATGATGAAACA 600

Qy     583 GACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCT 642
Db     601 ACCTTAGAACAAATCGAAAAAGACATTAAAGAACAAGTAATCAATGAGGTGATTCCTCAT 660

Qy     643 GAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGT 702
Db     661 GAACTTTTAGATGATGAAACAAAATACTTTATCAACCCAACTGGTCGCTTTGTAATCGGT 720

Qy     703 GGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGC 762
Db     721 GGCCCTCAAGGAGATGCCGGCTTAACTGGTCGTAAAATTATCGTTGATACTTATGGCGGT 780

Qy     763 ATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCC 822
Db     781 TATGCTCGTCATGGCGGCGGTGCTTTTTCTGGAAAAGATGCAACCAAAGTTGACCGCTCT 840

Qy     823 GCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGT 882
Db     841 GCTAGTTATGCTGCGCGTTATATCGCGAAAAACATTGTTGCTGCTGGTTTGGCGAAAAAA 900

Qy     883 TGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAA 942
Db     901 GTAGAAGTCCAATTAGCTTATGCAATTGGCGTGGCGCAACCAGTGTCTATTTCTATTAAT 960

Qy     943 ACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTC 1002
Db     961 ACGTTTGGTACGAGTGCTCTTCCAGAAAGTAAATTAATCGAAGCAGTTCGTAAAAACTTT 1020

Qy    1003 GACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAA 1062
Db    1021 GATTTACGTCCAGCAGGTATTATTGAAATGCTAGATTTACGTCGCCCAATTTATAAACAA 1080

Qy    1063 ACCGCAGCATACGGTCACTTTGGTCG------TGAACATTTCCCGTGGGAAAAAACCGAC 1116
Db    1081 ACTGCGGCTTATGGTCATTTTGGTCGAACAGATATTGATTTACCGTGGGAACAAACAGAT 1140

Qy    1117 AAAG 1120
Db    1141 AAAG 1144

Search completed: April 22, 2003, 15:13:25
Job time : 188.738 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:05:34 ; Search time 1882.93 Seconds
(without alignments)
9934.410 Million cell updates/sec

Title: US-09-622-419-1
Perfect score: 1155
Sequence: 1 atggcaaaacacctttttac.....atgctgccggtctgaagtaa 1155

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	271.4	23.5	1912	11	AY109333	AY109333 Zea mays
	2	248	21.5	994	14	BM807271	BM807271 AGENCOURT
	3	246.6	21.4	920	17	BH770810	BH770810 LLMGtag55
	4	235	20.3	889	17	CNS06M4T	AL405027 T7 end of
	5	233.2	20.2	1056	10	BE035235	BE035235 MO02F06 M
	6	232.8	20.2	1060	10	BE412603	BE412603 MCG002.B0
	7	225.2	19.5	1045	14	BQ071434	BQ071434 AGENCOURT
c	8	219.2	19.0	921	17	CNS070VU	AL424144 T7 end of
	9	219.2	19.0	1071	14	BQ055329	BQ055329 AGENCOURT
	10	214.4	18.6	977	14	BQ964578	BQ964578 AGENCOURT
	11	209.6	18.1	768	12	BF626331	BF626331 HVSMEa001
	12	209	18.1	755	13	BI921837	BI921837 EST541740
	13	207.6	18.0	817	12	BG414252	BG414252 HVSMEk000
c	14	207.4	18.0	720	13	BJ432712	BJ432712 BJ432712
c	15	207.4	18.0	725	13	BJ432117	BJ432117 BJ432117
c	16	207.4	18.0	743	13	BJ428430	BJ428430 BJ428430
c	17	207.4	18.0	750	13	BJ428525	BJ428525 BJ428525
	18	207	17.9	798	12	BF629512	BF629512 HVSMEb001
c	19	206.8	17.9	771	13	BJ373284	BJ373284 BJ373284
	20	206.6	17.9	858	14	BQ165469	BQ165469 EST611338
	21	206	17.8	914	14	BM815992	BM815992 HC109A08_
	22	204.6	17.7	692	12	BF480223	BF480223 L0-2124T3
c	23	204.6	17.7	808	17	BH404603	BH404603 AG-ND-153
	24	204.4	17.7	955	14	BQ950501	BQ950501 AGENCOURT
	25	204	17.7	738	13	BI423145	BI423145 EST533811
	26	202.2	17.5	850	12	BF259971	BF259971 HVSMEf002
c	27	202	17.5	631	13	BJ436098	BJ436098 BJ436098
	28	202	17.5	731	14	BU004375	BU004375 QGG5A11.y
c	29	201	17.4	618	13	BJ433832	BJ433832 BJ433832
c	30	201	17.4	645	13	BJ430220	BJ430220 BJ430220
c	31	201	17.4	646	13	BJ433905	BJ433905 BJ433905
c	32	201	17.4	648	13	BJ430052	BJ430052 BJ430052
c	33	201	17.4	687	13	BJ435244	BJ435244 BJ435244
	34	201	17.4	691	9	AU034101	AU034101 AU034101
c	35	201	17.4	695	13	BJ435499	BJ435499 BJ435499
c	36	201	17.4	698	13	BJ430606	BJ430606 BJ430606
c	37	201	17.4	699	13	BJ431424	BJ431424 BJ431424
c	38	201	17.4	699	13	BJ431971	BJ431971 BJ431971
c	39	201	17.4	701	13	BJ434216	BJ434216 BJ434216
c	40	201	17.4	713	13	BJ429678	BJ429678 BJ429678
	41	201	17.4	830	14	BQ096649	BQ096649 IfHdk0007
c	42	200.4	17.4	610	13	BJ429470	BJ429470 BJ429470
c	43	200.4	17.4	652	13	BJ373810	BJ373810 BJ373810
c	44	200.4	17.4	693	13	BJ374845	BJ374845 BJ374845
c	45	200	17.3	707	13	BJ432258	BJ432258 BJ432258

ALIGNMENTS

RESULT 1
AY109333/c
LOCUS AY109333 1912 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays CL2757_5 mRNA sequence.
ACCESSION AY109333
VERSION AY109333.1 GI:21212903
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 1912)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/DuPont Consensus Sequences for Design of Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 1912)
AUTHORS Coe,E.C.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
FEATURES Location/Qualifiers
source 1. .1912
/organism="Zea mays"
/db_xref="MaizeDB:631135"
/db_xref="taxon:4577"
/clone="CL2757_5"

```
                     /clone_lib="Maize Mapping Project/DuPont Cornsensus
                     Library"
                     /note="this sequence is part of a project of EST
                     assemblies resulting from the application of public
                     contigs to seed DuPont contigs; this resource was
                     assembled by DuPont as part of a collaboration for the
                     overgo addressing of BACs in conjunction with the Maize
                     Mapping Project"
BASE COUNT      361 a    509 c    504 g    390 t    148 others
ORIGIN

  Query Match           23.5%;  Score 271.4;  DB 11;  Length 1912;
  Best Local Similarity   54.8%;  Pred. No. 2.1e-71;
  Matches  612;  Conservative    0;  Mismatches  472;  Indels   33;  Gaps    4;

Qy       2 TGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTG 61
Db    1601 TCGACACATTCCTCTTCACCTCGGAGTCTGTGAACGAGGGACACCCTGACAAGCTCTGTG 1542

Qy      62 ACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTG 121
Db    1541 ACCAGGTCTCAGATGCCGTTCTTGACGCTTGCCTTGCTGAGGACCCTGACAGCAAGGTTG 1482

Qy     122 CTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCG 181
Db    1481 CTTGTGAGACCTGCACCAAGACCAACATGGTCATGGTGTTCGGCGAGATCACGACCAAGG 1422

Qy     182 CCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATT 241
Db    1421 CGACCGTGGACTACGAGAAGATTGTCAGGGAGACCTGCCGCAACATTGGTTTTGTGTCAA 1362

Qy     242 CCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTC 301
Db    1361 ACGATGTTGGGCTTGACGCCGACCACTGCAAGGTGCTCGTGAACATTGAGCAGCAGTCCC 1302

Qy     302 CTGACATCAACCAGGGCGTTGACCGTGCC---------GATCCGCTGGAACAGGGCGCGG 352
Db    1301 CTGATATTGCTCAGGGTGTGCATGGCCACTTCACCAAGCGCCCCGAGGAGATTGGAGCTG 1242

Qy     353 GTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCAC 412
Db    1241 GTGACCAGGGACACATGTTCGGGTATGCGACCGACGAGACCCCTGAGCTGATGCCCCTCA 1182

Qy     413 CTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTC 472
Db    1181 GCCATGTCCTTGCCACCAAGCTGGGCGCGCGTCTCACCGAGGTGCGCAAGAACGGAACCT 1122

Qy     473 TGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTA----------TG 520
Db    1121 GCCCCTGGCTCAGGCCCGATGGGAAGACCCAGGTGACAGTCGAGTACCGCAATGAGGGTG 1062

Qy     521 ACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGA 580
Db    1061 GCGCCATGGTCCCCATCCGTGTGCACACCGTCCTCATCTCCACCCAGCACGACGAGACAG 1002

Qy     581 TCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCG 640
Db    1001 TGACCAACGATGAGATCGCTGCTGACCTGAAGGAGCATGTCATCAAGCCTGTCATCCCTG 942

Qy     641 CTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCG 700
Db     941 AGCAGTACCTCGACGAGAAGACCATCTTCCACCTCAACCCGTCCGGGCGCTTCGTCATCG 882

Qy     701 GTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCG 760
Db     881 GCGGGCCCCACGGCGATGCCGGCCTCACCGGCCGCAA------CAACGACACGTACGGCG 828

Qy     761 GCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTT 820
Db     827 GCTGGGGAGCCCACGGCGGTGGCGCCTTCTCCGGCAAGGACCCCACCAAGGTGGACCGCA 768

Qy     821 CCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATC 880
```

```
Db     767 GCGGCGCCTACGTGGCCAAGGCAGGCCCCAAGAGCATCGTGGCCNNNNNNNNNNNNNNNN 708

Qy     881 GTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAG 940
Db     707 NNNNNNNNNNGCAGGTGTCGTACGCCATCGGCGTGCCGGAGCCCCTGTCCGTGTTCGTCA 648

Qy     941 AAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCT 1000
Db     647 ACTCGTACGGCACCGGCACGATCCCCGACAAGGAGATCCTCAAGATCGTGAAGGAGAACT 588

Qy    1001 TCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCA------CCCGATCT 1054
Db     587 TCGACTTCAGGCCCGGGATGATCAGCATCAACCTCGACCTGAAGAAGGGCGGCAACAGGT 528

Qy    1055 ACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGA 1091
Db     527 TCATCAAGACCGCCGCATACGGCCACTTTGGCCGTGA 491


RESULT 2
BM807271
LOCUS       BM807271                994 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6566487 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:5738319
            5', mRNA sequence.
ACCESSION   BM807271
VERSION     BM807271.1  GI:19124094
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 994)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12749  row: n  column: 16
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .994
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5738319"
                     /clone_lib="NIH_MGC_88"
                     /tissue_type="duodenal adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: small intestine; Vector: pCMV-SPORT6;
                     Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
                     oligo-dT primed. Average insert size 1.767 kb. Library
                     enriched for full-length clones and constructed by Life
                     Technologies. Note: this is a NIH_MGC Library."
BASE COUNT      257 a    198 c    250 g    285 t      4 others
ORIGIN

  Query Match           21.5%;  Score 248;  DB 14;  Length 994;
  Best Local Similarity   58.3%;  Pred. No. 2.1e-64;
  Matches  515;  Conservative    0;  Mismatches  351;  Indels   17;  Gaps    4;

Qy      46 CCTGACAAAATTGCTGACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGAT 105
Db       2 CCAGATAAGATTTGTGACCAAATCAGTGATGCTGTCCTTGATGCCCACCTTCAGCAGGAT 61

Qy     106 CCGAAAGCACGCGTTGCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGC 165
Db      62 CCTGATGCCAAAGTAGCTTGTGAAACTGTTGCTAAAACTGGAATGATCCTTCTTGCTGGG 121
```

```
Qy    166 GAAATCACCACCAGCGCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAA 225
Db    122 GAAATTACATCCAGAGCTGCTGTTGACTACCAGAAAGTGGTTCGTGAAGCTGTTAAACAC 181

Qy    226 ATTGGCTATGTGCATTCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCT 285
Db    182 ATTGGATATGATGATTCTTCCAAAGGTTTTGACTACAAGACTTGTAACGTGCTGGTAGCC 241

Qy    286 ATCGGCAAACAGTCTCCTGACATCAACCAGGGCGTTGACCGTGCC---GATCCGCTGGAA 342
Db    242 TTGGAGCAACAGTCACCAGATATTGCTCAAGGTGTTCATCTTGACAGAAATGAAGAAGAC 301

Qy    343 CAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTG 402
Db    302 ATTGGTGCTGGAGACCAGGGCTTAATGTTTGGCTATGCCACTGATGAAACTGAGGAGTGT 361

Qy    403 ATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAA 462
Db    362 ATGCCTTTAACCATTGTCTTGGCACACAAGCTAAATGCCAAACTGGCAGAACTACGCCGT 421

Qy    463 AACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGAC 522
Db    422 AATGGCACTTTGCCTTGGTTACGCCCTGATTCTAAAACTCAAGTTACTGTGCAGTATATG 481

Qy    523 GACGGCAAAATCGTTG-----------GTATCGATGCTGTCGTGCTTTCCACTCAGCAC 570
Db    482 CAGGATCGAGGTGCTGTGCTTCCCATCAGAGTCCACACAATTGTTATATCTGTTCAGCAT 541

Qy    571 TCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCA 630
Db    542 GATGAAGAGGTTTGTCTTGATGAAATGAGGGATGCCCTAAAGGAGAAAGTCATCAAAGCA 601

Qy    631 ATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGT 690
Db    602 GTTGTGCCTGCGAAATACCTTGATGAGGATACAATCTACCACCTACAGCCAAGTGGCAGA 661

Qy    691 TTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGAT 750
Db    662 TTTGTTATTGGTGGGCCTCAGGGTGATGCTGGTTTGACTGGACGGAAAATCATTGTGGAC 721

Qy    751 ACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGA-TCCATCAAA 809
Db    722 ACTTATGGCGGTTGGGGTGCTCATGGAGGAGGTGCCTTTTCAGGAAAGGATTATACCAAG 781

Qy    810 AGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGG 869
Db    782 GGTCGACCGTTCAGCTGCTTATGCTGCTCGTTGGGTGGCAAAATCCCTTGTTAAAGGGAG 841

Qy    870 CCTG-GCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGG 911
Db    842 TCTGTGCCGGGAGGGTTCTTGTTCANGTCTCTTATGCTATTGG 884
```

```
RESULT 3
BH770810
LOCUS       BH770810                920 bp    DNA     linear   GSS 01-MAY-2002
DEFINITION  LLMGtag552 MG1363 Random Sequence Tag Library Lactococcus lactis
            subsp. cremoris genomic, DNA sequence.
ACCESSION   BH770810
VERSION     BH770810.1  GI:20373767
KEYWORDS    GSS.
SOURCE      Lactococcus lactis subsp. cremoris.
  ORGANISM  Lactococcus lactis subsp. cremoris
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Lactococcus.
REFERENCE   1  (bases 1 to 920)
  AUTHORS   Bolotin,A., Ehrlich,S.D. and Sorokin,A.
  TITLE     Studies of genomes of dairy bacteria Lactococcus lactis
  JOURNAL   Sci. Aliments,  (2002) In press
COMMENT     Contact: Sorokin A
            Genetique Microbienne
            INRA
            CRJ INRA, Domaine de Vilvert, 78352 Jouy en Josas cedex, France
            Tel: 33 1 34 65 25 16
            Fax: 33 1 34 65 25 21
            Email: sorokine@jouy.inra.fr
            best homologue in strain IL1403 is metK (90%)
            Class: shotgun
            High quality sequence start: 30
            High quality sequence stop: 893.
FEATURES             Location/Qualifiers
     source          1..920
                     /organism="Lactococcus lactis subsp. cremoris"
                     /strain="MG1363"
                     /db_xref="taxon:1359"
                     /clone_lib="MG1363 Random Sequence Tag Library"
                     /note="Vector: pSGMU2; Site_1: SmaI; Library of
                     chromosomal fragments of L.lactis strain MG1363 was
                     prepared by partial AluI digestion or by sonication."
BASE COUNT      271 a    169 c    207 g    273 t
ORIGIN
```

Query Match 21.4%; Score 246.6; DB 17; Length 920;
Best Local Similarity 59.1%; Pred. No. 5.3e-64;
Matches 482; Conservative 0; Mismatches 319; Indels 14; Gaps 3;

```
Qy    346 GGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATG 405
Db     60 GGTGCTGGTGACCAAGGAATGATGTTCGGTTTTGCAACTAACGAAACTGCCGAATACATG 119

Qy    406 CCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAAC 465
Db    120 CCACTTGCCATTTCACTTTCACATAAATTAGTTAAGAAATTGGCTGATTTGCGTAAATCA 179

Qy    466 GGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATG---AC 522
Db    180 GGTGAAATTTCATACTTGCGTCCAGATGCAAAATCACAAGTGACTGTAGAATATGATAAT 239

Qy    523 GACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATC 582
Db    240 AATGGAAAAGCAAAACGTGTTGATACCGGTGTTATCTCGACTCAACATGCTGCTACGGCA 299

Qy    583 GACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCT 642
Db    300 ACAAATGAAGAAATTCATGATGATGTGATTAACAAAGTGATTAAGGCTATTATTGCTGCC 359

Qy    643 GAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGT 702
Db    360 GAACTTTTGGATGATGAAACATAATACTTTATCAATCCAACATGACGTTTTGTTATCGGT 419

Qy    703 GGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGC 762
Db    420 GGTCCTCAAGGAGACTCAGGTTTGACTGGCCGTAAGATTATTGTTGATACTAACGGGGGG 479

Qy    763 ATGGCGCGTCACGGTGGCGGT-----GCATTCTCTGGTAAAGATCCATCAAAAGTGGACC 817
Db    480 GTTATGCACCACATGAGTGGTTGGGGCGCTTCTCAGGTAAGGATGCAACTAAGGTTGACC 539

Qy    818 GTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCG 877
Db    540 GTTCAGCTTCTTACGCGGCTCGTTATGTCGCAAAAAACATTGTGGCTGCTGGACTTGCTG 599

Qy    878 ATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGG 937
Db    600 ATAAAGCTCAAATTCAACTTTCTTATGCGATTGGTGTGGCCACTCCTACATCAATCAATG 659

Qy    938 TAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGT 997
Db    660 TTGACACTTTTGGTACAGGTAAAGTTTCAGATAATGAACTTAGGTCAACGATACGCAAGG 719

Qy    998 TCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACA 1057
Db    720 TCTATCTACTCCGTGTATCACGAATTATCGTAAGGCTTGACGTCCTTCGTATAAATCATG 779

Qy   1058 AAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACAT------TTCCCGTGGGAAAAAA 1111
```

```
Db    780 TACAGACAACAGCTTATAGGCACTTCGGCCGTACGGATATTGAATTGCCATGGGAACAAA 839

Qy   1112 CCGACAAAGCGCAGCTGCTGCGCGATGCTGCCGGT 1146
Db    840 CAGACAAGGTTGAAGAATTGAAAGCTGTTTTAGGT 874
```

RESULT 4
CNS06M4T
LOCUS CNS06M4T 889 bp DNA linear GSS 17-JUN-2001
DEFINITION T7 end of clone AU0AA002A05 of library AU0AA from strain CBS 3082 of Saccharomyces kluyveri, genomic survey sequence.
ACCESSION AL405027
VERSION AL405027.1 GI:12167244
KEYWORDS GSS.
SOURCE Saccharomyces kluyveri.
ORGANISM Saccharomyces kluyveri
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 889)
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 889)
AUTHORS Neuveglise,C., Bon,E., Lepingle,A., Wincker,P., Artiguenave,F., Gaillardin,C. and Casaregola,S.
TITLE Genomic exploration of the hemiascomycetous yeasts: 9. Saccharomyces kluyveri
JOURNAL FEBS Lett. 487 (1), 56-60 (2000)
MEDLINE 20584719
PUBMED 11152884
REFERENCE 3 (bases 1 to 889)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (07-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES Location/Qualifiers
source 1. .889
/organism="Saccharomyces kluyveri"
/strain="CBS 3082"
/db_xref="taxon:4934"
/clone="AU0AA002A05"
/clone_lib="AU0AA"
/note="end : T7"
misc_feature <1. .>888
/note="similar to Saccharomyces cerevisiae ORF YDR502c [SAM2 ; S-adenosylmethionine synthetase 2] similar to Saccharomyces cerevisiae ORF YLR180w [SAM1 ; S-adenosylmethionine synthetase 1]"
/evidence=not_experimental
BASE COUNT 214 a 229 c 226 g 197 t 23 others
ORIGIN

Query Match 20.3%; Score 235; DB 17; Length 889;
Best Local Similarity 54.7%; Pred. No. 1.9e-60;
Matches 481; Conservative 16; Mismatches 368; Indels 15; Gaps 2;

```
Qy    241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
Db     10 TCCGATAAGGGTTTCGACTACAAGACCTGTAACGTGTTGGTTGCTGTCGAACAACAGTCT 69

Qy    301 CCTGACATCAACCAGGGCGTTGACCGTGCCGATCCGCTGGAACAG---GGCGCGGGTGAC 357
Db     70 CCAGACATTGCTCAGGGTTTGCACTACGAACAAGCGCTAGAAGACCTAGGTGCCGGTGAC 129

Qy    358 CAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATC 417
Db    130 CAAGGTATCATGTTTGGKTASGCCACCGACGAAASCCCAGAAGGTTTGCCGTTGACCATT 189

Qy    418 ACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCG 477
Db    190 TTGTTGGCCCACAAGCTGAACATCGCCATGGCCGACGCCAGAAGAGACGGCTCTTTGGSA 249

Qy    478 TGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACG---------- 526
Db    250 TGGTTGAGASSTGASASCAAGASSCAGGTSACCGTSGAGTACAAGGAAGASCASGGTAGA 309

Qy    527 -GCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGAC 585
Db    310 TGGATTSCATTGAGAATTGACACCGTTGTTGTGTCTGSCCAACACGCCGACGAGATCTCC 369

Qy    586 CAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAA 645
Db    370 ACCGAAGACTTGAGATCTTTGATCAAGTCCGAGATCATCCAAAAGGTCATCBCAGCCGAC 429

Qy    646 TGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGC 705
Db    430 ATGTTGGACGAAAAGACCAAGTACTACATCCAGCCTTCCGGCAGATTCGTGATTGGTGGT 489

Qy    706 CCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATG 765
Db    490 CCTCAAGGTGACGCCGGTTTGACCGGTAGAAAGATCATCGTCGACGCCTACGGTGGTGCC 549

Qy    766 GCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCA 825
Db    550 TCTGCCGTCGGTGGTGGTGCCTTCTCCGGTAAGGACTACTCCAAGGTCGANCGTTCCGCT 609

Qy    826 GCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGT 885
Db    610 GCCTACGCCGCCAGATGGGTGGGCAAGTCCTTGGTGCACSCCSGTTTSTSCAAGAGAGTC 669

Qy    886 GAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACT 945
Db    670 CAAGTGCAATTCTCCTACGCCATCGGTATTGCCGAGCCTCTGTCCTTGCACGTCGACACC 729

Qy    946 TTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGTGAGTTCTTCGAC 1005
Db    730 TACGGTACCGCCACCAAGTCCGACGAGGAAATCATCGAGATCATCAAGAAGAACTTTGAC 789

Qy   1006 CTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATCTACAAAGAAACC 1065
Db    790 TTGAGACCAGGTGTCTTGGTCAAGGAGTTGGACTTGGCTAGACCAATCTACTTGCCAACT 849

Qy   1066 GCAGCATACGGTCACTTTGGTCGTGAACATTTCCCGTGGG 1105
Db    850 GCTTCCTACGGTCACTTTACCAACCASGAATACCCATGGG 889
```

RESULT 5
BE035235
LOCUS BE035235 1056 bp mRNA linear EST 07-JUN-2000
DEFINITION MO02F06.MO Mesembryanthemum crystallinum cDNA 5' similar to s-adenosylmethionine synthetase #, mRNA sequence.
ACCESSION BE035235
VERSION BE035235.1 GI:8330359
KEYWORDS EST.

```
SOURCE      common iceplant.
  ORGANISM  Mesembryanthemum crystallinum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Caryophyllidae; Caryophyllales; Aizoaceae; Mesembryanthemum.
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Bohnert,H.J., Borchert,C., Brazille,S., Brooks,J., Eaton,M., Ferrea
            ,H., Kawasaki,S., McCollough,A., Michalowski,C.B., Palacio,C.,
            Scara,G., Wheeler,M. and Zepeda,G.R.
  TITLE     Functional Genomics of Plant Stress Tolerance
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Michalowski,C.B.
            University of Arizona
            Bio Sciences West room 513, Tucson, AZ 85721, USA
            Tel: 520-621-7982
            Fax: 520-621-1697
            Email: cbm@u.arizona.edu
            An open reading frame exists.
            Insert Length: 1   Std Error: 0.00.
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Mesembryanthemum crystallinum"
                     /db_xref="taxon:3544"
                     /clone_lib="MO"
                     /tissue_type="apical meristem and leaf primordia"
                     /dev_stage="5 weeks"
                     /note="no stress"
BASE COUNT      229 a    298 c    264 g    263 t      2 others
ORIGIN

  Query Match             20.2%;  Score 233.2;  DB 10;  Length 1056;
  Best Local Similarity   56.9%;  Pred. No. 7.3e-60;
  Matches  497;  Conservative    0;  Mismatches  355;  Indels   22;  Gaps    3;

Qy       1 ATGGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCT 60
Db     133 ATGGACACCTTCCTATTCACCTCTGAATCCGTCAACGAGGGGCACCCCGACAAGCTCTGC 192

Qy      61 GACCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTT 120
Db     193 GATCAGATCTCTGATGCAGTTCTTGATGCTTGCCTTGCGCAAGATGCTGAGAGCAAGGTT 252

Qy     121 GCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGC 180
Db     253 GCTTGTGAGACTTGTACCAAGACCAACTTGGTCATGGTTTTCGGTGAGATCACCACCAAG 312

Qy     181 GCCTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCAT 240
Db     313 GCCAATGTAGACTATGAGAAGATTGCCCGTGACACTTGCCGCAACATTGGGTTCGTGTCC 372

Qy     241 TCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCT 300
Db     373 GATGATGTTGGTCTTGATGCCGACAACTGCAAGGTCTTGGTCTACATTGAGCAACAGAGC 432

Qy     301 CCTGACATCAACCAGGGCGTTGACCGTGCCGAT---------CCGCTGGAACAGGGCGCG 351
Db     433 CCTGACATTGCCCAGGGTGTTCATGGTCACCTTACCAAGCGCCCTGAGGAGATTGGTGCT 492

Qy     352 GGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCA 411
Db     493 GGTGACCAGGGACACATGTTTGGCTATGCCACTGATGAGACCCCCGAGTTGATGCCCCTT 552

Qy     412 CCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACCGGCACT 471
Db     553 AGCCATGTCCTTGCTACCAAGCTCGGTGCCCGCTTGACCGAGGTCCGTAAGAACGGAACC 612

Qy     472 CTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAA 531
Db     613 TGCCCGTGGTTGAGGCCTGATGGAAAGACCCAAGTCACTGTTGAGTACTACAACGAANAC 672

Qy     532 ATC------------GTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAG 579
Db     673 GGTGCCATGGTTCCTGTCCGTGTCCACACTGTGCTCATCTCAACCCAGCACGATGAGACC 732

Qy     580 ATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCC 639
Db     733 GTCACCAACGACGAGAATGCTGCTGATTTGAAGGAGCACGTCATCAAGCCCGTCTTCCCC 792

Qy     640 GCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGAC-CGGTCGTTTCGTTAT 698
Db     793 GAGAAGTACTTCGATGAGAAGACCATCTTNCACCTCAACCCATCTTGCCCGGTTCGTGAA 852

Qy     699 CGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGG 758
Db     853 TGGTGGACCCCATGGTGACGCTGGTCTCACTGGCCGTAAGATCATCAATCATACCTATGG 912

Qy     759 CGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCG 818
Db     913 TGGGTTGGGGGCCCACGGGGGTGGTGCTTTCTTCGAAAAGGACCAAACCCAGGTCAACAG 972

Qy     819 TTCCGCAGCCTACGCAGCACGTTATGTCGCGAAA 852
Db     973 GAGTGTGGCTTACATTCAAAGAAAAGCCCTAAAA 1006


RESULT 6
BE412603
LOCUS       BE412603                1060 bp    mRNA    linear   EST 24-JUL-2000
DEFINITION  MCG002.B05R990625 ITEC MCG Barley Leaf/Culm Library Hordeum vulgare
            cDNA clone MCG002.B05, mRNA sequence.
ACCESSION   BE412603
VERSION     BE412603.1  GI:9410449
KEYWORDS    EST.
SOURCE      Hordeum vulgare.
  ORGANISM  Hordeum vulgare
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Hordeum.
REFERENCE   1  (bases 1 to 1060)
  AUTHORS   Anderson,O.A., Appels,R., Bailey,P., Blake,T., Close,T., Cloutier
            ,S., Dubcovsky,J., Feuillet,C., Gale,M., Graner,A., Gustafson,P.,
            Herrmann,R.G., Holton,T., Jacquemin,J.M., Jia,J., Joudrier,P.,
            Langridge,P., Lazo,G.R., Lin,J.J., McGuire,P., Ogihara,Y.,
            Pecchioni,N., Qualset,C., Schuch,W., Selvaraj,G., Shariflou,M.,
            Sorrells,M., Warburton,M. and Wenzel,G.
  TITLE     International Triticeae EST Cooperative (ITEC): Production of
            Expressed Sequence Tags for Species of the Triticeae
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Graner A
            Institute for Plant Genetics & Crop Plant Research
            Corrensstr. 3, D-06466 Gatersleben GERMANY
            Tel: 49 39482 5521
            Fax: 49 39482 5137
            Email: a_graner@ipk-gatersleben.de
            International Triticeae EST Cooperative (ITEC)
            http://wheat.pw.usda.gov/genome.
FEATURES             Location/Qualifiers
     source          1. .1060
                     /organism="Hordeum vulgare"
                     /db_xref="taxon:4513"
                     /clone="MCG002.B05"
                     /clone_lib="ITEC MCG Barley Leaf/Culm Library"
                     /tissue_type="leaf/culm"
                     /dev_stage="etiolated"
BASE COUNT      231 a    332 c    288 g    198 t     11 others
ORIGIN

  Query Match             20.2%;  Score 232.8;  DB 10;  Length 1060;
  Best Local Similarity   59.0%;  Pred. No. 9.7e-60;
  Matches  480;  Conservative    0;  Mismatches  311;  Indels   23;  Gaps    4;

Qy      12 CCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTC 71
Db     145 CCTCTTCACTTCCGAGTCCGTGAACGAGGGCCATCCCGACAAGTTGTGTGACCAGGTCTC 204

Qy      72 TGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAAC 131
```

```
Db    205 CGACGCCGTCCTGGACGCCTGCTTGGCCCAGGATCCTGACAGCAAGGTTGCCTGCGAGAC 264

Qy    132 CTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGA 191

Db    265 CTGCACCAAGACCAACATGGTCATGGTCTTCGGTGAGATCACCACCAAGGCCACCGTGGA 324

Qy    192 CATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATGGG 251

Db    325 CTATGAGAAGATCGTGCGTGACACCTGCCGCGACATTGGCTTCATCTCTGACGACGTCGG 384

Qy    252 CTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAA 311

Db    385 TCTCGACGCCGACCATTGCAAGGTGCTCGTCAACATCGAGCAACAATCCCCTGACATTGC 444

Qy    312 CCAGGGCGTTGACCGTGCC----------GATCCGCTGGAACAGGGCGCGGGTGACCAGGG 362

Db    445 CCAGGGTGTTCATGGACACTTCACCAAGCGTCCAGAAGAGGTCGGCGCCGGTGACCAGGG 504

Qy    363 TCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTA 422

Db    505 CATCATGTTTGGCTATGCCACGGATGAGACCCCTGAGCTGATGCCCCTCACCCACATGCT 564

Qy    423 TGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCT 482

Db    565 TGCCACCAAGCTCGGAGCTCGCCTCACCGAGGTCCGCAAGAATGGCACCTGTGCCTGGCT 624

Qy    483 GCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTA----------TGACGACGGCAA 530

Db    625 GAGGCCTGATGGAAAGACCCAGGTCACCATTGAGTACCTAAATGAGGGTGGTGCCATGGT 684

Qy    531 AATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAA 590

Db    685 GCCTGTTCGTGTGCACACCGTNCTCATCTCCACCCAGCATGACGAGACCGTCACCAACGA 744

Qy    591 ATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAA-TGGC 649

Db    745 TGAGATCGCCGCGGACCTCAAGGAGCACGTCATCAAGCCGGTGATTCCCGGGAAAGTATC 804

Qy    650 TGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAA 709

Db    805 TCGATGAAAACACCATCTTCCACCTGAACCCGTCTGGTCGCTTCGTCATCGGCGGCCCTC 864

Qy    710 TGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGC 769

Db    865 ATGGCGATGCCGGTCT-ACCGGCCGCAAGATCATCATNGACACCTACGGTGGCTGGGGAG 923

Qy    770 GTCACGGTGGCGGTGCATTCTCTGGTAAAGATCC 803

Db    924 CCCATGGCGGCGGNGCTTTNTCTGGCAAGGACCC 957


RESULT 7
BQ071434
LOCUS       BQ071434                1045 bp    mRNA    linear   EST 02-APR-2002
DEFINITION  AGENCOURT_6867291 NIH_MGC_47 Homo sapiens cDNA clone IMAGE:5927842
            5', mRNA sequence.
ACCESSION   BQ071434
VERSION     BQ071434.1  GI:19900480
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1045)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2103  row: g  column: 11
            High quality sequence stop: 670.
FEATURES             Location/Qualifiers
     source          1..1045
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5927842"
                     /clone_lib="NIH_MGC_47"
                     /tissue_type="neuroblastoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: brain; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming.  Directionally
                     cloned into EcoRI/XhoI sites using the following 5'
                     adaptor: GGCACGAG(G).  Size-selected >500bp for average
                     insert size 1.8kb.  Library constructed by Ling Hong in
                     the laboratory of Gerald M. Rubin (University of
                     California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
BASE COUNT      252 a    218 c    270 g    303 t      2 others
ORIGIN

  Query Match           19.5%;  Score 225.2;  DB 14;  Length 1045;
  Best Local Similarity   56.6%;  Pred. No. 2e-57;
  Matches  444;  Conservative    0;  Mismatches  329;  Indels   12;  Gaps    1;

Qy    346 GGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATG 405

Db      7 GGTGCTGGAGACCAGGGCTTAATGTTTGGCTATGCCACTGATGAAACTGAGGAGTGTATG 66

Qy    406 CCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAAC 465

Db     67 CCTTTAACCATTGTCTTGGCACACAAGCTAAATGCCAAACTGGCAGAACTACGCCGTAAT 126

Qy    466 GGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGAC 525

Db    127 GGCACTTTGCCTTGGTTACGCCCTGATTCTAAAACTCAAGTTACTGTGCAGTATATGCAG 186

Qy    526 GGCAAAATCGTTG----------GTATCGATGCTGTCGTGCTTTCCACTCAGCACTCT 573

Db    187 GATCGAGGTGCTGTGCTTCCCATCAGAGTCCACACAATTGTTATATCTGTTCAGCATGAT 246

Qy    574 GAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATT 633

Db    247 GAAGAGGTTTGTCTTGATGAAATGAGGGATGCCCTAAAGGAGAAAGTCATCAAAGCAGTT 306

Qy    634 CTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTC 693

Db    307 GTGCCTGCGAAATACCTTGATGAGGATACAATCTACCACCTACAGCCAAGTGGCAGATTT 366

Qy    694 GTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACC 753

Db    367 GTTATTGGTGGGCCTCAGGGTGATGCTGGTTTGACTGGACGGAAAATCATTGTGGACACT 426

Qy    754 TACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTG 813

Db    427 TATGGCGGTTGGGGTGCTCATGGAGGAGGTGCCTTTTCAGGAAAGGATTATACCAAGGTC 486

Qy    814 GACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTG 873

Db    487 GACCGTTCAGCTGCTTATGCTGCTCGTTGGGTGGCAAAATCCCTTGTTAAAGGAGGTCTG 546

Qy    874 GCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATC 933

Db    547 TGCCGGAGGGTTCTTGTTCAGGTCTCTTATGCTATTGGAGTTTCTCATCCATTATCTATC 606

Qy    934 ATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAACTGACCCTGCTGGTACGT 993

Db    607 TCCATTTTCCATTATGGTACCTCTCAGAAGAGTGAGAGAGAGCTATTAGAGATTGTGAAG 666
```

```
Qy     994 GAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGGATCTGCTGCACCCGATC 1053
           ||   ||||| || |||||    |   |  |   |  |||||||||  | | || ||
Db     667 AAGAATTTCGATCTCCGCCCTGNGGTCATTGTCAGGGATCTGGATCTGAAGAAGCCAATT 726

Qy    1054 TACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACATTTCCCGTGGGAAAAAACC 1113
           ||  |    || ||||| || || |||||||||   ||    ||||| ||||||    ||
Db     727 TATCAGAGGACTGCAGCCTATGGCCACTTTGGTAAGGACAGCTTCCCATGGGAAGTGCCC 786

Qy    1114 GACAA 1118
           | ||
Db     787 AAAAA 791


RESULT 8
CNS070VU/c
LOCUS       CNS070VU                921 bp    DNA     linear   GSS 06-JUL-2001
DEFINITION  T7 end of clone AZ0AA012F03 of library AZ0AA from strain CBS 712 of
            Kluyveromyces marxianus, genomic survey sequence.
ACCESSION   AL424144
VERSION     AL424144.1  GI:12207338
KEYWORDS    GSS.
SOURCE      Kluyveromyces marxianus.
  ORGANISM  Kluyveromyces marxianus
            Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
            Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
REFERENCE   1  (bases 1 to 921)
  AUTHORS   Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
            Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
            de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
            Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
            Saurin,W., Tekaia,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
            Wincker,P. and Weissenbach,J.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 1. A set of
            yeast species for molecular evolution studies
  JOURNAL   FEBS Lett. 487 (1), 3-12 (2000)
  MEDLINE   20584711
   PUBMED   11152876
REFERENCE   2  (bases 1 to 921)
  AUTHORS   Llorente,B., Malpertuy,A., Blandin,G., Artiguenave,F., Wincker,P.
            and Dujon,B.
  TITLE     Genomic exploration of the hemiascomycetous yeasts: 12.
            Kluyveromyces marxianus var. marxianus
  JOURNAL   FEBS Lett. 487 (1), 71-75 (2000)
  MEDLINE   20584722
   PUBMED   11152887
REFERENCE   3  (bases 1 to 921)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
            2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
            seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT     This GSS is part of a random genomic sequencing program of thirteen
            yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
            exiguus, Saccharomyces servazzii, Zygosaccharomyces rouxii,
            Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
            lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
            angusta, Debaryomyces hansenii var.  hansenii, Pichia sorbitophila,
            Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
            5 kb were prepared and both extremities were sequenced. See
            keywords for description of this sequence and for the sequence of
            the other extremity of this insert.
FEATURES             Location/Qualifiers
     source          1. .921
                     /organism="Kluyveromyces marxianus"
                     /strain="CBS 712"
                     /variety="marxianus"
                     /db_xref="taxon:4911"
                     /clone="AZ0AA012F03"
                     /clone_lib="AZ0AA"
                     /note="end : T7"
     misc_feature    complement(<41. .>921)
                     /note="similar to Saccharomyces cerevisiae ORF YDR502c [
                     SAM2 ; S-adenosylmethionine synthetase 2 ]
                     2 putative frameshift(s)
                     similar to Saccharomyces cerevisiae ORF YLR180w [ SAM1 ;
                     S-adenosylmethionine synthetase 1 ]"
                     /evidence=not_experimental
BASE COUNT      225 a    223 c    233 g    238 t      2 others
ORIGIN

  Query Match           19.0%;  Score 219.2;  DB 17;  Length 921;
  Best Local Similarity   56.8%;  Pred. No. 1.3e-55;
  Matches  506;  Conservative    1;  Mismatches  365;  Indels   19;  Gaps    5;

Qy      70 TCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAA 129
           ||||| ||  | || || ||    ||   | |||| ||     |    |||||||| |||
Db     921 TCTGACGCTATCTTGGATGCTTGTCTAAGAGAGGACCCATTGTCCAAGGTTGCTTGTGAA 862

Qy     130 ACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTA 189
           ||    |  || ||||| ||| || | |||  ||| |||||||||||||| |||    |
Db     861 ACTGCAGCCAAGACCGGTATGATTATGGTTTTCGGTGAAATCACCACCAGAGCCAACTTG 802

Qy     190 GACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATG 249
           |||  | || ||||    ||| ||||  |     |||| || || |    |||||| | |
Db     801 GACTACCAAAAGATTGTGCGTGACACGATCAAGCAAATCGGTTACGACTCTTCCGAGAAG 742

Qy     250 GGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATC 309
           || || |||   ||  | |||   ||| ||   || || |  |||| || || || |||
Db     741 GGTTTCGACTACAAGACATGTAACGTTTTGGTTGCCATTGAACAACAATCCCCAGATATC 682

Qy     310 AACCAGGGCGTTGACCGTGCCGATCCGCTGGAACA---GGGCGCGGGTGACCAGGGTCTG 366
           || ||  |  ||   |  | ||||| |  |||| |   |||| |||||||| ||| |
Db     681 GCTCAAGGTTTGCACTACGAACAAGCTTTGGAAGAATTGGGTGCTGGTGACCAAGGTATC 622

Qy     367 ATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCA 426
           ||||| || ||||| ||| | ||||||   |  |  |||||   | ||    |  ||
Db     621 ATGTTCGGTTACGCTACTGACGAAACCCCAGAACACTTGCCATTGACCATTTTGTTGGCC 562

Qy     427 CACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGC 486
           |||    ||  |    | |   ||| |   | |  |||| || |||| ||| |  |
Db     561 CACAAGTTG--AACATGGCCATGTGACGCCAGAAGGGACGGTTCTTTGCCATGGATCAGA 504

Qy     487 CCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTGGTATC--- 543
           || ||| | || | ||| || ||  |  | || | || | | |         | |||
Db     503 CCAGACACCAAGACCCAAGTCACCATCGAATACAAGGAAGACCACGGTAGATGGATCCCA 444

Qy     544 ---------GATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCG 594
                   ||  |  ||||  | ||  | || ||| | || || |||  |
Db     443 TTGAGAATTGACACCATCGTCATCTCTGCCCAACACGCCGAGGAAATCTCCACCGCTGAC 384

Qy     595 CTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACT 654
           |||   || |  |||      ||||||||| |     |    || ||||  | ||
Db     383 TTGCGTCAATCCCTAAGAACCGAGATCATCGACAAGGTCTGCCCAGCTGAGATGTTGGAC 324

Qy     655 TCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGT 714
             |||||| | || |||| | ||  |||| | |||| |||||||| |||   |||
Db     323 GAAAACACCAAGTACTACATCCAACCATCCGGTAGATTCGTCATCGGTGGTCCACAAGGT 264

Qy     715 GACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCAC 774
           |||   ||| |||| ||| | || || ||||| |||  |||||| ||     |   |
Db     263 GACGCTGGTTTGACCGGTAGAAAGATCATCGTCGAT-GCTACGGTGGTGCCTCCGCTGTT 205

Qy     775 GGTGGCGGTGCATTCTCTGGTAAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCA 834
           ||||| ||||| ||||| ||||| |||    || || || |||||||| ||  : |||||
Db     204 GGTGGTGGTGCTTTCTCCGGTAAGGATTACTCCAAGGTCGACCGTTCTGC-TSTTACGCT 146

Qy     835 GCACGTTATGTCGCGAAAAACATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAG 894
           ||  | |  ||||| ||    | | ||     ||||  |      || |    || | ||
Db     145 GCTAGATGGGTCGCCAAGTCCTTGGTCCACGCTGGTATCTGTAAGAGAGTGCAAGTCCAA 86

Qy     895 GTTTCCTACGCAATCGGCGTGGCTGAACCGACCTCCATCATGGTAGAAACT 945
           | || ||||| || ||  | ||||| ||     ||  |    |||| | ||
Db      85 TTCTCTTACGCTATTGGTATCGCTGAGCCATTGTCTTTGCACGTTGNACCT 35
```

```
RESULT 9
BQ055329
LOCUS       BQ055329                1071 bp    mRNA    linear   EST 29-MAR-2002
DEFINITION  AGENCOURT_6797446 NIH_MGC_99 Homo sapiens cDNA clone IMAGE:5807491
            5', mRNA sequence.
ACCESSION   BQ055329
VERSION     BQ055329.1  GI:19814669
KEYWORDS    EST.
SOURCE      human..
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1071)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lou Staudt
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2049  row: p  column: 20
            High quality sequence stop: 714.
FEATURES             Location/Qualifiers
     source          1. .1071
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5807491"
                     /clone_lib="NIH_MGC_99"
                     /tissue_type="lymphoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: lymph; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Size-selected >500bp for average insert size
                     1.8kb. Library constructed by Ling Hong in the laboratory
                     of Gerald M. Rubin (University of California, Berkeley)
                     using ZAP-cDNA synthesis kit (Stratagene) and Superscript
                     II RT (Life Technologies). Note: this is a NIH_MGC
                     Library."
BASE COUNT      279 a    238 c    263 g    287 t      4 others
ORIGIN

  Query Match             19.0%;  Score 219.2;  DB 14;  Length 1071;
  Best Local Similarity   57.9%;  Pred. No. 1.4e-55;
  Matches  453;  Conservative    0;  Mismatches  314;  Indels   16;  Gaps    3;

Qy       3 GGCAAAACACCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGA 62
           ||  | |  |||||| || || ||||| |||   ||||| || || || || |||  |||
Db      29 GGGCACATTCCTTTTCACCTCAGAGTCGGTCGGGGAAGGCCACCCAGATAAGATTTGTGA 88

Qy      63 CCAAATTTCTGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGC 122
           ||||||   |||||| ||  | || ||   |||  | ||||||||  | ||     || ||
Db      89 CCAAATCAGTGATGCTGTCCTTGATGCCCACCTTCAGCAGGATCCTGATGCCAAAGTAGC 148

Qy     123 TTGCGAAACCTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGC 182
           ||| |||||    |  ||||| || ||| |   |  |||  || ||||| ||  |||| ||
Db     149 TTGTGAAACTGTTGCTAAAACTGGAATGATCCTTCTTGCTGGGGAAATTACATCCAGAGC 208

Qy     183 CTGGGTAGACATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTC 242
             || |||  | |  |  |    ||| |  | |||    | ||||| ||||   ||||
Db     209 TGCTGTTGACTACCAGAAAGTGGTTCGTGAAGCTGTTAAACACATTGGATATGATGATTC 268

Qy     243 CGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCC 302
             ||  || ||||||   ||  | |||   || |||   ||  | |   |||||| ||
Db     269 TTCCAAAGGTTTTGACTACAAGACTTGTAACGTGCTGGTAGCCTTGGAGCAACAGTCACC 328

Qy     303 TGACATCAACCAGGGCGTTGACCGTGCC---GATCCGCTGGAACAGGGCGCGGGTGACCA 359
             || ||    || || ||| | | || |    ||      ||    || || || |||||

Db     329 AGATATTGCTCAAGGTGTTCATCTTGACAGAAATGAAGAAGACATTGGTGCTGGAGACCA 388

Qy     360 GGGTCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCAC 419
           |||  | ||||||||||| || ||| |||||| || | |  |||||   | | ||
Db     389 GGGCTTAATGTTTGGCTATGCCACTGATGAAACTGAGGAGTGTATGCCTTTAACCATTGT 448

Qy     420 CTATGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTG 479
           ||  ||||||   ||          | ||| ||| | ||   || |||||| |||| ||
Db     449 CTTGGCACACAAGCTAAATGCCAAACTGGCAGAACTACGCCGTAATGGCACTTTGCCTTG 508

Qy     480 GCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTATGACGACGGCAAAATCGTTG- 538
           | | |||||  ||  | ||||  || || ||| | ||||||  | |  |   | | ||
Db     509 GTTACGCCCTGATTCTAAAACTCAAGTTACTGTGCAGTATATGCAGGATCGAGGTGCTGT 568

Qy     539 ----------GTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCA 587
                      |  || |  |  | ||  | ||   ||||||   |||||||  |    |
Db     569 GCTTCCCATCAGAGTCCACACAATTGTTATATCTGTTCAGCATGATGAAGAGGTTTGTCT 628

Qy     588 GAAATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATG 647
             |    ||  || ||  ||| |||  |  |||||||  || || |||| ||  |||
Db     629 TGATGAAATGAGGGATGCCCTAAAGGAGAAAGTCATCAAAGCAGTTGTGCCTGCGAAATA 688

Qy     648 GCTGACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCC 707
            ||  |  |  || |  | |  | |  | || |  ||  | || ||||| ||||| ||
Db     689 CCTTGATGAGGATACAATCTACCACCTACAGCCAAGTGGCAGATTTGTTATTGGTGGGCC 748

Qy     708 AATGGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCTACGGCGGCATGGC 767
             |||||    ||| |||||| ||  |||| || || || || || ||  |  |||
Db     749 TCANGGTGATGCTGGTTTGACTGGACG-CAAATCATTGTGGACACTTATGGGCGGTTGGG 807

Qy     768 GCG 770
           | |
Db     808 GTG 810


RESULT 10
BQ964578
LOCUS       BQ964578                977 bp    mRNA    linear   EST 21-AUG-2002
DEFINITION  AGENCOURT_10050435 NIH_MGC_134 Mus musculus cDNA clone
            IMAGE:6514516 5', mRNA sequence.
ACCESSION   BQ964578
VERSION     BQ964578.1  GI:22380056
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 977)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. David Rowe
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14089  row: p  column: 05
            High quality sequence start: 213
            High quality sequence stop: 405.
FEATURES             Location/Qualifiers
     source          1. .977
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:6514516"
                     /clone_lib="NIH_MGC_134"
                     /tissue_type="undifferentiated limb"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Vector: pCMV-SPORT6.1.ccdb; Site_1: EcoRV; Site_2:
                     NotI; Cloned unidirectionally. Primer: Oligo dT. Average
```

insert size 1.7 kb. Constructed by ResGen, Invitrogen Corp. Note: this is a NIH_MGC Library."
BASE COUNT 252 a 190 c 256 g 276 t 3 others
ORIGIN

Query Match 18.6%; Score 214.4; DB 14; Length 977;
Best Local Similarity 56.0%; Pred. No. 4e-54;
Matches 492; Conservative 0; Mismatches 369; Indels 18; Gaps 4;

```
Qy    271 GCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAACCAGGGCGTTGACCGTGCC 330
Db     21 GGGATGTTGGTTGCCTTNGGACAACAGTCACCAGATATTGCCCAAGGTGTTCATCTTGAC 80

Qy    331 ---GATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAAT 387
Db     81 CGGAATGAGGAAGATATTGGTGCAGGAGACCAGNGTTTGATGTTTGGTTATGCCACTGAT 140

Qy    388 GAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAG 447
Db    141 GAAACTGAAGAGTGTATGCCTTTAACCATTGTCTTAGCACACAAGCTAAATGCCAAATTG 200

Qy    448 GCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTG 507
Db    201 GCTGAACTACGCCGCAATGGTACATTGCCTTGGTTACGCCCAGATTCTAAAACTCAAGTG 260

Qy    508 ACTTTTCAGTAT----------GACGACGGCAAAATCGTTGGTATCGATGCTGTCGTG 555
Db    261 ACTGTGCAATATATGCAAGATCGGGGTGCTGTGCTTCCCATCAGAGTCCACACGATTGTT 320

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615
Db    321 ATATCTGTTCAGCACGATGAAGAAGTTTGTCTTGATGAGATGAGGGATGCTCTGAAGGAG 380

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATC 675
Db    381 AAAGTGATCAAGGCTGTTGTACCTGCAAAATACCTTGATGAGGATACAATTTACCACCTA 440

Qy    676 AACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGT 735
Db    441 CAGCCAAGTGGCAGATTTGTTATTGGTGGGCCTCAGGGTGATGCTGGCCTGACTGGCCGA 500

Qy    736 AAAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGT 795
Db    501 AAAATCATTGTGGATACTTATGGCGGTTGGGGAGCTCATGGAGGAGGGGCCTTTTCAGGA 560

Qy    796 AAAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAAC 855
Db    561 AAGGATTATACCAAAGTGGACCGTTCAGCTGCTTATGCTGCTCGTTGGGTGGCAAAATCC 620

Qy    856 ATCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTG 915
Db    621 CTTGTTAAAGGAGGTCTGTGCAGGAGGGTTCTTGTTCAGGTCTCTTATGCTATTGGAGTT 680

Qy    916 GCTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAA 975
Db    681 TCTCATCCATTGTCGATCTCCATTTTCCATTATGGCACTTCTCAGAAGAGTGAGAGAGAG 740

Qy    976 CTGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCA-GATGCT 1034
Db    741 CTATTAGAAATTGTAAGGAAGAATTTTGATCTTCGCCCTGNGGTCATTGGTCAGGGATCT 800

Qy   1035 GGATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACA 1094
Db    801 GGATCTGAAGAAACCAATTTATCAGAGGACTGCAGCCTATGGCCACCTTTGGTAAGGACA 860

Qy   1095 --TTTCCCGTGGGAAAAAACCGACAAAGCGCAGCTGCTG 1131
Db    861 GCTTTCCCATGGGAAATGCCCCAAAAAGCTTAAATATTG 899
```

RESULT 11
BF626331
LOCUS BF626331 768 bp mRNA linear EST 18-OCT-2001
DEFINITION HVSMEa0018G02f Hordeum vulgare seedling shoot EST library HVcDNA0001 (Cold stress) Hordeum vulgare cDNA clone HVSMEa0018G02f, mRNA sequence.
ACCESSION BF626331
VERSION BF626331.3 GI:16256799
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 768)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R., Choi,D.W., Fenton,R.D. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex cold-stressed seedling shoot cDNA library
JOURNAL Unpublished (2001)
COMMENT On Dec 18, 2000 this sequence version replaced gi:13086209.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 580
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 735.
FEATURES Location/Qualifiers
source 1. .768
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEa0018G02f"
/clone_lib="Hordeum vulgare seedling shoot EST library HVcDNA0001 (Cold stress)"
/tissue_type="Seedling shoot"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1; Seeds were surface sterilized then germinated under axenic conditions in the dark at room temperature on filter paper with water, nystatin and cefotaxime in covered crystallization dishes. Five-day old seedlings were incubated at 5oC for 2 days. Shoots were then harvested, total RNA was prepared, poly(A) RNA was purified, one primary unamplified cDNA library was made, and 600000 pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids. These steps were performed in the TJ Close laboratory at the University of California, Riverside (Choi, Close, Fenton). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 162 a 250 c 210 g 145 t 1 others
ORIGIN

Query Match 18.1%; Score 209.6; DB 12; Length 768;
Best Local Similarity 57.1%; Pred. No. 1e-52;
Matches 433; Conservative 0; Mismatches 304; Indels 21; Gaps 2;

```
Qy    223 GAAATTGGCTATGTGCATTCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTTCTGAGC 282
          || |||||||   |   |  |||| | ||  | |||||  ||   ||   ||| ||   |
Db      1 GACATTGGCTTCATCTCTGACGACGTCGGTCTCGACGCCGACCATTGCAAGGTGCTCGTC 60

Qy    283 GCTATCGGCAAACAGTCTCCTGACATCAACCAGGGCGTTGACCGTGCCGA---------T 333
          ||||   |||| || ||||||||   |||||| ||| |  |   |          |
Db     61 AACATCGAGCAACAATCCCCTGACATTGCCCAGGGTGTTCATGGACACTTCACCAAGCGT 120

Qy    334 CCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAATGAAACC 393
          ||    ||   ||||| |||||||||||  | ||||||||||| || ||  |||| |||
Db    121 CCAGAAGAGGTCGGCGCCGGTGACCAGGGCATCATGTTTGGCTATGCCACGGATGAGACC 180

Qy    394 GACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAGGCTGAA 453
          | |||||||||      |   ||    |||   |   || | |   || |    | ||
Db    181 CCTGAGCTGATGCCCCTCACCCACATGCTTGCCACCAAGCTCGGAGCTCGCCTCACCGAG 240

Qy    454 GTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGACTTTT 513
          || || || || |||||     | |||||| | || || |  || | ||||||  ||  ||
Db    241 GTCCGCAAGAATGGCACCTGTGCCTGGCTGAGGCCTGATGGAAAGACCCAGGTCACCATT 300

Qy    514 CAGTA------------TGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGCTTTCC 561
          ||||            ||  | |          ||| || |  |  | ||| |  | |||
Db    301 GAGTACCTAAATGAGGGTGGTGCCATGGTGCCTGTTCGTGTGCACACCGTCCTCATCTCC 360

Qy    562 ACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAGAGATC 621
          || |||||    ||     ||  | |  |   | |    |  |   | | |||  |  ||
Db    361 ACCCAGCATGACGAGACCGTCACCAACGATGAGATCGCCGCGGACCTCAAGGAGCACGTC 420

Qy    622 ATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCAACCCG 681
          ||||||||  |  | ||||   | |  ||   |     |||||  |||  | | ||||||
Db    421 ATCAAGCCGGTGATTCCCGGGAAGTACCTCGATGAAAACACCATCTTCCACCTGAACCCG 480

Qy    682 ACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTAAAATT 741
          | ||||| |||||  ||||| |||||    || ||   ||||||  || || || || || ||
Db    481 TCTGGTCGCTTCGTCATCGGCGGCCCTCATGGCGATGCCGGTCTCACCGGCCGCAAGATC 540

Qy    742 ATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTAAAGAT 801
          |||  | || |||||||| |||  ||      || || ||||| || |||||||| || ||
Db    541 ATCATCGACACCTACGGTGGCTGGGGAGCCCATGGCGGCGGCGCTTTCTCTGGCAAGGAC 600

Qy    802 CCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACATCGTT 861
          ||  | || || |||||    | ||||||   ||  |  | | ||| || |  |||| |
Db    601 CCGACCAAGGTCGACCGCAGTGGCGCCTACATTGCCAGGCAAGCCGCCAAGAGCATCATC 660

Qy    862 GCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGGCTGAA 921
          ||    || || ||    || ||      | ||| | || ||||||||||| ||  |||||
Db    661 GCCAGCGGTCTCGCACGCCGCTGCATTGTGCAGATCTCATACGCAATCGGGGTACCTGAA 720

Qy    922 CCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAA 959
          ||        |  | || ||  | | ||| || | |||
Db    721 CCCTTGATTGTGTTCGTCGACTCCTACGGGACCGGGAA 758


RESULT 12
BI921837
LOCUS       BI921837                 755 bp    mRNA    linear   EST 17-OCT-2001
DEFINITION  EST541740 tomato callus Lycopersicon esculentum cDNA clone
            cLEC74D16 5' end, mRNA sequence.
ACCESSION   BI921837
VERSION     BI921837.1  GI:16217865
KEYWORDS    EST.
SOURCE      tomato.
  ORGANISM  Lycopersicon esculentum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum;
            Lycopersicon.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Alcala,J., Vrebalov,J., White,R., Vision,T., Karamycheva,S.A., Tsai
            ,J., Utterback,T., Van Aken,S., Ronning,C.M., Fraser,C.M., Martin
            ,G.B., Tanksley,S.D. and Giovannoni,J.
  TITLE     Generation of ESTs from tomato callus tissue (2001)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: CUGI
            Clemson University Genomics Institute
            Clemson University
            100 Jordan Hall, Clemson, SC 29634, USA
            Email: http://www.genome.clemson.edu/orders/index.html
            This clone is available through the Clemson University Genomics
            Institute
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Lycopersicon esculentum"
                     /cultivar="TA496"
                     /db_xref="taxon:4081"
                     /clone="cLEC74D16"
                     /clone_lib="tomato callus"
                     /tissue_type="callus"
                     /dev_stage="25-40 days old"
                     /lab_host="XL1-Blue MRF'"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Boyce Thompson Institute; sequencing: The
                     Institute for Genomic Research; cLEC - Cotyledons of
                     seedlings 7-10 days post-germination were excised, cut at
                     both ends and placed on MS medium with no selection. Mixed
                     callus was harvested at 25 and 40 days and included
                     undifferentiated masses. Tomato Callus EST Library"
BASE COUNT       210 a    196 c    184 g    165 t
ORIGIN

  Query Match           18.1%;  Score 209;  DB 13;  Length 755;
  Best Local Similarity  57.3%;  Pred. No. 1.5e-52;
  Matches  430;  Conservative    0;  Mismatches  300;  Indels   21;  Gaps    2;

Qy     37 GAAGGGCATCCTGACAAAATTGCTGACCAAATTTCTGATGCCGTTTTAGACGCGATCCTC 96
          ||||| || || ||||| |  |||||| | || |||||| || | || ||   |||
Db      2 GAAGGTCACCCCGACAAGCTCTGTGACCAGGTCTCAGATGCCATTCTTGATGCTTGCCTC 61

Qy     97 GAACAGGATCCGAAAGCACGCGTTGCTTGCGAAACCTACGTAAAAACCGGCATGGTTTTA 156
          || ||||||||  ||      ||||| || ||||||| |  ||| ||   ||||||| |
Db     62 GAGCAGGATCCCGAAAGCAAGGTTGCATGTGAAACCTGCACAAAGACAAACATGGTTATG 121

Qy    157 GTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGACATCGAAGAGATCACCCGTAACACC 216
          ||    || || ||||| ||||  |||  ||||||  ||| | |||||   ||| ||||
Db    122 GTCTTTGGAGAGATCACAACCAAGGCCACCGTAGACTACGAGAAGATCGTACGTGACACA 181

Qy    217 GTTCGCGAAATTGGCTATGTGCATTCCGACATGGGCTTTGACGCTAACTCCTGTGCGGTT 276
             | |  ||||| |  ||     | ||  | ||  |||| ||| ||  |||   |||
Db    182 TGCAGAGGCATTGGGTTCGTCTCAGCAGATGTTGGTCTTGATGCTGACAACTGCAAGGTC 241

Qy    277 CTGAGCGCTATCGGCAAACAGTCTCCTGACATCAACCAGGGCGTTGACCGT--------- 327
          ||       || |   |||||   |||||||||  ||| || ||| || ||
Db    242 CTTGTTAACATTGAACAACAGAGCCCTGACATCGCCCAAGGAGTTCACGGTCATCTTACC 301

Qy    328 GCCGATCCGCTGGAACAGGGCGCGGGTGACCAGGGTCTGATGTTTGGCTACGCAACTAAT 387
             | ||    |||   || || |||||||| ||||  |||||||| || || ||  ||
Db    302 AAGAAACCAGAAGAAATTGGAGCTGGTGACCAAGGTCATATGTTTGGGTATGCCACAGAT 361

Qy    388 GAAACCGACGTGCTGATGCCAGCACCTATCACCTATGCACACCGTCTGGTACAGCGTCAG 447
          |||||    | |||  |||||      |       |  ||  |   || |        |
Db    362 GAAACTCCTGAGCTCATGCCCCTCACCCATGTTTTGGCCACCAAGCTTGGTGCCAAGCTT 421

Qy    448 GCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTG 507
          ||||||||  | ||  |||   ||    || |||||  | ||  ||  |  ||  | ||| ||
Db    422 ACTGAAGTGAGGAAGAACAAAACCTGCCCATGGCTCAGGCCCGATGGCAAGACCCAAGTT 481

Qy    508 ACTTTTCAGTATGACGACGGCAAAATCG-----------TTGGTATCGATGCTGTCGTG 555
          ||| || ||||  |  ||| |||   |            ||  |  |  ||||  |
Db    482 ACTGTTGAGTACAAGAACGACAATGGTGCCATGGTCCCTATTAGAGTTCACACTGTTCTC 541

Qy    556 CTTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAA 615
          | || || ||||    |||    ||  | |  |   | |    |  |  | ||  | ||
Db    542 ATCTCAACCCAGCATGACGAAACTGTCACCAACGACCAGATCGCCCAGGACTTGAAAGAG 601

Qy    616 GAGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATC 675
```

```
Db    602 CATGTGATCAAGCCTGTGATCCCAGCTAAGTACCTTGACGAGAACACCATCTTCCACCTC 661

Qy    676 AACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGT 735

Db    662 AACCCATCAGGTCGCTTCGTCATTGGTGGTCCACATGGAGATGCTGGTCTTACTGGCAGG 721

Qy    736 AAAATTATCGTTGATACCTACGGCGGCATGG 766

Db    722 AAAATCATCATTGATACCTACGGGAGCTGGG 752
```

RESULT 13
BG414252
LOCUS BG414252 817 bp mRNA linear EST 23-OCT-2001
DEFINITION HVSMEk0002D12f Hordeum vulgare testa/pericarp EST library HVcDNA0013 (normal) Hordeum vulgare cDNA clone HVSMEk0002D12f, mRNA sequence.
ACCESSION BG414252
VERSION BG414252.1 GI:13319899
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 817)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Kannangara,G., von Wetstein,D., Akhunov,E., Chin,A., Choi,D.W., Fenton,R.D., Kianian,P., Otto,C., Simons,K., Zhang,D., Begum,D., Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource for barley genomics: Morex testa/pericarp cDNA library
JOURNAL Unpublished (2001)
COMMENT Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hq bases = 347
Seq primer: AATTAACCCTCACTAAAGGG
High quality sequence stop: 643.
FEATURES Location/Qualifiers
source 1. .817
/organism="Hordeum vulgare"
/cultivar="Morex"
/db_xref="taxon:4513"
/clone="HVSMEk0002D12f"
/clone_lib="Hordeum vulgare testa/pericarp EST library HVcDNA0013 (normal)"
/tissue_type="testa/pericarp"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xhol; Plants were raised from seeds in a Controlled Environments growth chamber maintained in continuous light at 18oC, and testa and pericarp were dissected from developing kernels at Washington State University, Pullman, WA (Kannangara, von Wetstein). Total RNA was prepared, poly(A) RNA was purified, one cDNA library was made, and 1 million pfu were in vivo excised to give pBluescript SK(-) cDNA phagemids in the TJ Close lab at the University of California, Riverside (Akhunov, Chin, Choi, Close, Fenton, Kianian, Otto, Simons, Zhang). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"
BASE COUNT 182 a 256 c 219 g 160 t
ORIGIN

Query Match 18.0%; Score 207.6; DB 12; Length 817;
Best Local Similarity 56.8%; Pred. No. 4.3e-52;
Matches 434; Conservative 0; Mismatches 309; Indels 21; Gaps 2;

```
Qy     12 CCTTTTTACGTCCGAGTCCGTCTCTGAAGGGCATCCTGACAAAATTGCTGACCAAATTTC 71

Db     53 CCTCTTCACTTCCGAGTGGGTGAACGAGGGCCATCCCGACAAGTTGTGTGACCAGGTCTC 112

Qy     72 TGATGCCGTTTTAGACGCGATCCTCGAACAGGATCCGAAAGCACGCGTTGCTTGCGAAAC 131

Db    113 CGACGCCGTGCTGGACGCCTGCTTGGCCCAGGATCCTGACAGCAAGGTTGTCTGCGAGAC 172

Qy    132 CTACGTAAAAACCGGCATGGTTTTAGTTGGCGGCGAAATCACCACCAGCGCCTGGGTAGA 191

Db    173 CTGCACCAAGACCAACATGGTCATGGTCTTCGGTGAGATCACCACCAAGGCCACCGTGGA 232

Qy    192 CATCGAAGAGATCACCCGTAACACCGTTCGCGAAATTGGCTATGTGCATTCCGACATGGG 251

Db    233 CTATGAGAAGATCGTGCGTGACACCTGCCGCGACATTGGCTTCATCTCTGACGACGTCGG 292

Qy    252 CTTTGACGCTAACTCCTGTGCGGTTCTGAGCGCTATCGGCAAACAGTCTCCTGACATCAA 311

Db    293 TCTCGACGCCGACCATTGCAAGGTGCTCGTCAACATCGAGCAACAATCCGCTGACATTGC 352

Qy    312 CCAGGGCGTTGACCGTGCC---------GATCCGCTGGAACAGGGCGCGGGTGACCAGGG 362

Db    353 CCAGGGTGTTCATGGACACTGTACCAAGCGTCCAGAAGAGGTCGGCGCCGGTGACCAGGG 412

Qy    363 TCTGATGTTTGGCTACGCAACTAATGAAACCGACGTGCTGATGCCAGCACCTATCACCTA 422

Db    413 CATCATGTTTGGCTATGCCACGGATGAGACCCCTGAGCTGATGCCCCTCACCCACATGCT 472

Qy    423 TGCACACCGTCTGGTACAGCGTCAGGCTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCT 482

Db    473 TGCCACCAAGCTCGGAGCTCGCCTAACCGAGGTCCGCAAGAATGGCACCTGTGCCTGGCT 532

Qy    483 GCGCCCGGACGCGAAAAGCCAGGTGACTTTTCAGTA------------TGACGACGGCAA 530

Db    533 GAGGCCTGATGGACAGACCCAGGTCACCATTGAGTACCTAAATGAGGGTGGTGCCATGGT 592

Qy    531 AATCGTTGGTATCGATGCTGTCGTGCTTTCCACTCAGCACTCTGAAGAGATCGACCAGAA 590

Db    593 GCCTGTTCGTGTGCACACCGTCCTCATCTTCACCCAGCATGACGAGACCGTCACCAACGA 652

Qy    591 ATCGCTGCAAGAAGCGGTAATGGAAGAGATCATCAAGCCAATTCTGCCCGCTGAATGGCT 650

Db    653 TGAGATCGCCGCGGACCTCAAGGAGCACGTCTTCAAGCCGGTGATTCCCGGGAAGTACCT 712

Qy    651 GACTTCTGCCACCAAATTCTTCATCAACCCGACCGGTCGTTTCGTTATCGGTGGCCCAAT 710

Db    713 CGATGAAAACACCATTTTTCACCTGAACCCGTCTGGTCGCTTCCGTATCGGCGGCCCTCA 772

Qy    711 GGGTGACTGCGGTCTGACTGGTCGTAAAATTATCGTTGATACCT 754

Db    773 TGGCGATGCCCGGTCTACCGGCCGAAGAATCATCTTCGACACCT 816
```

RESULT 14
BJ432712/c
LOCUS BJ432712 720 bp mRNA linear EST 13-MAR-2002
DEFINITION BJ432712 Dictyostelium discoideum cDNA library, VF Dictyostelium discoideum cDNA clone ddv19c17 3', mRNA sequence.
ACCESSION BJ432712
VERSION BJ432712.1 GI:19407434

```
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 720)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .720
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv19c17"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT      210 a    149 c    157 g    203 t      1 others
ORIGIN

  Query Match             18.0%;  Score 207.4;  DB 13;  Length 720;
  Best Local Similarity   58.6%;  Pred. No. 4.6e-52;
  Matches  386;  Conservative    0;  Mismatches  261;  Indels   12;  Gaps    1;

Qy     449 CTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGA 508
           ||||| | ||| | || || ||  |  | |||   || || || || || |  || || |
Db     708 CTGAATTACGTCACAATGGTACCTTAGCCTGGGCTCGTCCAGATGCCAAGACTCAAGTCA 649

Qy     509 CTTTTCAGTA------------TGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGC 556
           || || | ||             | || |   ||     ||  |  |  |  |  ||||
Db     648 CTGTTGAATACAAAAAAGAAGGTCACAAAATCACTCCAGTCAGAGTACACACCATCGTTA 589

Qy     557 TTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAG 616
           |||| ||||| |||   ||| |  || | ||    |       |  |||||  | |    | |
Db     588 TTTCAACTCAACACGACGAAAAAGTCACCAACGAACAAATCAGAGCTGATTTGAAAGAAT 529

Qy     617 AGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCA 676
           |||||||  |  || | || ||| |||   |       |  |||||  | |    | |
Db     528 TAGTCATCAAAGCTGTTGTTCCAGCTCAATACTTAGACGATAACACCATCTACCATTTGA 469

Qy     677 ACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTA 736
           | ||  | ||||||||||| || ||||||||||||||||| |  ||| | || |||||||
Db     468 ATCCATCAGGTCGTTTCGTCATTGGTGGCCCAATGGGTGATTCTGGTTTAACCGGTCGTA 409

Qy     737 AAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTA 796
           | || ||| ||||| | || || ||   ||    |||||||||| ||||| ||||||||||
Db     408 AGATCATCATTGATTCATATGGTGGTTGGGGTGCTCACGGTGGTGGTGCTTTCTCTGGTA 349

Qy     797 AAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACA 856
           ||||| |  | ||||| || ||||| |  || |||||  ||  | |   |||| |||
Db     348 AAGATGCCACTAAAGTCGATCGTTCTGGTGCTTACGCTGCTAGATGGATCGCTAAATCTT 289

Qy     857 TCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGG 916
           | ||||| ||||| || ||||||||||||  | |||| ||||| ||| | || || || |
Db     288 TAGTTGCCGCTGGTCTTGCCGATCGTTGTTTAGTTCAAGTTTCATACTCTATTGGTGTTG 229

Qy     917 CTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAAC 976
           |  | ||   |   || | || || || |  ||||| |  |||  | ||  ||   ||
Db     228 CCAAGCCACTCAGTGTCTTTGTTGATACCTATGGTACCGGTAAAAAGACTGATGCTGAAA 169

Qy     977 TGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGG 1036
           |        |  |   | |   ||| ||  | || |||    || | || |     ||
Db     168 TCCTTGCCATCATCGATAATAACTTTGATTTACGTCCAGGTTGTTTAATGCGTGACCTCC 109

Qy    1037 ATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACAT 1095
```

```
           |  |    |  || || | | || |||| ||  |    |||||||||||| |  |  ||
Db     108 AATTAACTCGTCCAATTTTCCAAAAAACTGCCTCTGGTGGTCACTTTGGTAGAAATGAT 50


RESULT 15
BJ432117/c
LOCUS       BJ432117                725 bp    mRNA    linear   EST 13-MAR-2002
DEFINITION  BJ432117 Dictyostelium discoideum cDNA library, VF Dictyostelium
            discoideum cDNA clone ddv17o15 3', mRNA sequence.
ACCESSION   BJ432117
VERSION     BJ432117.1  GI:19406839
KEYWORDS    EST.
SOURCE      Dictyostelium discoideum.
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
  TITLE     Full length cDNA of Dictyostelium discoideum at the vegetative
            stage
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Dictyostelium discoideum"
                     /strain="AX4"
                     /db_xref="taxon:44689"
                     /clone="ddv17o15"
                     /clone_lib="Dictyostelium discoideum cDNA library, VF"
                     /sex="mat A"
                     /dev_stage="Growth phase"
BASE COUNT      211 a    150 c    156 g    207 t      1 others
ORIGIN

  Query Match             18.0%;  Score 207.4;  DB 13;  Length 725;
  Best Local Similarity   58.6%;  Pred. No. 4.6e-52;
  Matches  386;  Conservative    0;  Mismatches  261;  Indels   12;  Gaps    1;

Qy     449 CTGAAGTGCGTAAAAACGGCACTCTGCCGTGGCTGCGCCCGGACGCGAAAAGCCAGGTGA 508
           ||||| | ||| | || || ||  |  | |||   || || || || || |  || || |
Db     725 CTGAATTACGTCACAATGGTACCTTAGCCTGGGCTCGTCCAGATGCCAAGACTCAAGTCA 666

Qy     509 CTTTTCAGTA------------TGACGACGGCAAAATCGTTGGTATCGATGCTGTCGTGC 556
           || || | ||             | || |   ||     ||  |  |  |  |  ||||
Db     665 CTGTTGAATACAAAAAAGAAGGTCACAAAATCACTCCAGTCAGAGTACACACCATCGTTA 606

Qy     557 TTTCCACTCAGCACTCTGAAGAGATCGACCAGAAATCGCTGCAAGAAGCGGTAATGGAAG 616
           |||| ||||| |||   ||| |  || | ||    |       |  |||||  | |    | |
Db     605 TTTCAACTCAACACGACGAAAAAGTCACCAACGAACAAATCAGAGCTGATTTGAAAGAAT 546

Qy     617 AGATCATCAAGCCAATTCTGCCCGCTGAATGGCTGACTTCTGCCACCAAATTCTTCATCA 676
           |||||||  |  || | || ||| |||   |       |  |||||  | |    | |
Db     545 TAGTCATCAAAGCTGTTGTTCCAGCTCAATACTTAGACGATAACACCATCTACCATTTGA 486

Qy     677 ACCCGACCGGTCGTTTCGTTATCGGTGGCCCAATGGGTGACTGCGGTCTGACTGGTCGTA 736
           | ||  | ||||||||||| || ||||||||||||||||| |  ||| | || |||||||
Db     485 ATCCATCAGGTCGTTTCGTCATTGGTGGCCCAATGGGTGATTCTGGTTTAACCGGTCGTA 426

Qy     737 AAATTATCGTTGATACCTACGGCGGCATGGCGCGTCACGGTGGCGGTGCATTCTCTGGTA 796
           | || ||| ||||| | || || ||   ||    |||||||||| ||||| ||||||||||
Db     425 AGATCATCATTGATTCATATGGTGGTTGGGGTGCTCACGGTGGTGGTGCTTTCTCTGGTA 366

Qy     797 AAGATCCATCAAAAGTGGACCGTTCCGCAGCCTACGCAGCACGTTATGTCGCGAAAAACA 856
           ||||| |  | ||||| || ||||| |  || |||||  ||  | |   |||| |||
Db     365 AAGATGCCACTAAAGTCGATCGTTCTGGTGCTTACGCTGCTAGATGGATCGCTAAATCTT 306

Qy     857 TCGTTGCTGCTGGCCTGGCCGATCGTTGTGAAATTCAGGTTTCCTACGCAATCGGCGTGG 916
           | ||||| ||||| || ||||||||||||  | |||| ||||| ||| | || || || |
```

```
Db     305 TAGTTGCCGCTGGTCTTGCCGATCGTTGTTTAGTTCAAGTTTCATACTCTATTGGTGTTG 246

Qy     917 CTGAACCGACCTCCATCATGGTAGAAACTTTCGGTACTGAGAAAGTGCCTTCTGAACAAC 976
           |  | ||   |    || | || || || |  ||||| |  |||  | ||  ||   ||
Db     245 CCAAGCCACTCAGTGTCTTTGTTGATACCTATGGTACCGGTAAAAAGACTGATGCTGAAA 186

Qy     977 TGACCCTGCTGGTACGTGAGTTCTTCGACCTGCGCCCATACGGTCTGATTCAGATGCTGG 1036
           |        |  |   | |   ||| ||  | || |||    || | || |     ||
Db     185 TCCTTGCCATCATCGATAATAACTTTGATTTACGTCCAGGTTGTTTAATGCGTGACCTCC 126

Qy    1037 ATCTGCTGCACCCGATCTACAAAGAAACCGCAGCATACGGTCACTTTGGTCGTGAACAT 1095
           |  |    |  || || | | || |||| ||  |    ||||||||||| |  |  ||
Db     125 AATTAACTCGTCCAATTTTCCAAAAAACTGCCTCTGGTGGTCACTTTGGTAGAAATGAT 67

Search completed: April 22, 2003, 13:05:28
Job time : 1904.93 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:03:55 ; Search time 3355.45 Seconds
(without alignments)
10459.998 Million cell updates/sec

Title: US-09-622-419-3
Perfect score: 1206
Sequence: 1 atgaacgtttttaatcccgc.....tggaattattggtggattaa 1206

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		1206	100.0	1206	1	AB055109	AB055109 Escherich
2		1206	100.0	1206	6	AX018962	AX018962 Sequence
3		1206	100.0	3465	6	AX000478	AX000478 Sequence
4		1206	100.0	3794	6	AX000474	AX000474 Sequence
5		1206	100.0	3794	6	AX018970	AX018970 Sequence
6		1206	100.0	4975	6	AX018972	AX018972 Sequence
7		1206	100.0	12058	1	AE000364	AE000364 Escherich
8		1206	100.0	71128	1	ECU29581	U29581 Escherichia
9		1205	99.9	1216	6	AX000470	AX000470 Sequence
10		1167.6	96.8	266658	1	AP002563	AP002563 Escherich
11		1166.8	96.7	6488	1	AE005509	AE005509 Escherich
12	c	939	77.9	251891	2	AC084742	AC084742 Mus muscu
13		833.2	69.1	20653	1	AE008836	AE008836 Salmonell
14		823.6	68.3	230050	1	AL627277	AL627277 Salmonell
15		514.8	42.7	9280	1	AE013916	AE013916 Yersinia
16	c	514.8	42.7	210050	1	AJ414146	AJ414146 Yersinia
17		443	36.7	11773	2	AC023612	AC023612 Mus muscu
18		335.4	27.8	10786	1	AE004301	AE004301 Vibrio ch
19	c	268.6	22.3	2816	1	AB024603	AB024603 Burkholde
20	c	268.6	22.3	12537	1	AB024602	AB024602 Pseudomon
21		259.4	21.5	6866	1	ECH301654	AJ301654 Erwinia c
22		258.6	21.4	347660	1	AP002994	AP002994 Mesorhizo
23		257.2	21.3	10235	1	AE004786	AE004786 Pseudomon
24	c	252	20.9	11674	1	AE005391	AE005391 Escherich
25	c	252	20.9	257071	1	AP002558	AP002558 Escherich
26	c	251	20.8	217420	1	BSUB0017	Z99120 Bacillus su
27		244	20.2	1221	1	AB055108	AB055108 Escherich
28		244	20.2	1221	6	AX018966	AX018966 Sequence
29	c	244	20.2	11684	1	AE000263	AE000263 Escherich
30		244	20.2	13823	1	D90812	D90812 E.coli geno
31		244	20.2	19521	1	D90811	D90811 E.coli geno
32	c	242.2	20.1	14807	1	AB024601	AB024601 Pseudomon
33		237.8	19.7	13391	1	AE012389	AE012389 Xanthomon
34		237.2	19.7	10530	1	AE013796	AE013796 Yersinia
35		234.6	19.5	1217	6	AX436324	AX436324 Sequence
36		234.6	19.5	1224	6	AX436323	AX436323 Sequence
37	c	234	19.4	11137	1	AE005860	AE005860 Caulobact
38	c	225	18.7	4160	6	AX416825	AX416825 Sequence
39	c	224.6	18.6	14857	1	AE009543	AE009543 Brucella
40	c	221.2	18.3	40281	1	MLCB22	Z98741 Mycobacteri
41		221.2	18.3	348450	1	MLEPRTN4	AL583920 Mycobacte
42	c	220.8	18.3	15057	1	AE003977	AE003977 Xylella f
43	c	217.4	18.0	11206	1	AE009136	AE009136 Agrobacte
44	c	217.4	18.0	12426	1	AE008102	AE008102 Agrobacte
45		217	18.0	9988	1	AE011935	AE011935 Xanthomon

ALIGNMENTS

RESULT 1
AB055109
LOCUS AB055109 1206 bp DNA linear BCT 14-APR-2001
DEFINITION Escherichia coli csdA gene for NifS-like cysteine sulfinate desulfinase, complete cds.
ACCESSION AB055109
VERSION AB055109.1 GI:12619309
KEYWORDS .
SOURCE Escherichia coli (strain:JM109) DNA.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (sites)
AUTHORS Mihara,H., Kurihara,T., Yoshimura,T., Soda,K. and Esaki,N.
TITLE Cysteine sulfinate desulfinase, a NIFS-like protein of Escherichia

```
          coli with selenocysteine lyase and cysteine desulfurase activities.
          Gene cloning, purification, and characterization of a novel
          pyridoxal enzyme
  JOURNAL   J. Biol. Chem. 272 (36), 22417-22424 (1997)
  MEDLINE   97426379
REFERENCE   2  (sites)
  AUTHORS   Mihara,H., Kurihara,T., Yoshimura,T. and Esaki,N.
  TITLE     Kinetic and mutational studies of three NifS homologs from
          Escherichia coli: mechanistic difference between L-cysteine
          desulfurase and L-selenocysteine lyase reactions
  JOURNAL   J. Biochem. 127 (4), 559-567 (2000)
  MEDLINE   20206733
REFERENCE   3  (sites)
  AUTHORS   Lacourciere,G.M., Mihara,H., Kurihara,T., Esaki,N. and
          Stadtman,T.C.
  TITLE     Escherichia coli NifS-like proteins provide selenium in the pathway
          for the biosynthesis of selenophosphate
  JOURNAL   J. Biol. Chem. 275 (31), 23769-23773 (2000)
  MEDLINE   20379012
REFERENCE   4  (bases 1 to 1206)
  AUTHORS   Mihara,H., Kurihara,T. and Esaki,N.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-JAN-2001) Hisaaki Mihara, Institute for Chemical
          Research, Kyoto University, Laboratory of Molecular Microbial
          Science; Gokasho, Uji, Kyoto 611-0011, Japan
          (E-mail:mihara@mbc.kuicr.kyoto-u.ac.jp,
          URL:http://www.kuicr.kyoto-u.ac.jp/labos/bm2/lab.html,
          Tel:81-774-38-3243, Fax:81-774-38-3248)
FEATURES             Location/Qualifiers
     source          1. .1206
                     /organism="Escherichia coli"
                     /strain="JM109"
                     /db_xref="taxon:562"
                     /map="63.4 min"
     gene            1. .1206
                     /gene="csdA"
     CDS             1. .1206
                     /gene="csdA"
                     /function="Probably involved in the metabolism of
                     sulfur/selenium-containing amino acids"
                     /note="CSD, cysteine sulfinate desulfinase
                     Catalyzes the desulfination of L-cysteine sulfinate. Also
                     catalyzes the release of elemental sulfur and elemental
                     selenium from L-cysteine and L-selenocysteine,
                     respectively."
                     /codon_start=1
                     /transl_table=11
                     /product="NifS-like cysteine sulfinate desulfinase"
                     /protein_id="BAB21541.1"
                     /db_xref="GI:12619310"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT      256 a    315 c    359 g    276 t
ORIGIN

  Query Match           100.0%;  Score 1206;  DB 1;  Length 1206;
  Best Local Similarity  100.0%;  Pred. No. 1e-301;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200

Qy    1201 GATTAA 1206
```

```
          ||||||
Db   1201 GATTAA 1206


RESULT 2
AX018962
LOCUS       AX018962               1206 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 3 from Patent WO9942591.
ACCESSION   AX018962
VERSION     AX018962.1  GI:10043057
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1206)
  AUTHORS   Schroeder,H.
  TITLE     Method for producing biotin
  JOURNAL   Patent: WO 9942591-A 3 26-AUG-1999;
            SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES             Location/Qualifiers
     source          1. .1206
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             1. .1206
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC07715.1"
                     /db_xref="GI:10043058"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT      256 a    315 c    359 g    276 t
ORIGIN

  Query Match           100.0%;  Score 1206;  DB 6;  Length 1206;
  Best Local Similarity  100.0%;  Pred. No. 1e-301;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGCGG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200

Qy    1201 GATTAA 1206
           ||||||
Db    1201 GATTAA 1206


RESULT 3
AX000478
LOCUS       AX000478               3465 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 9 from Patent WO9905285.
ACCESSION   AX000478
VERSION     AX000478.1  GI:7240887
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 3465)
  AUTHORS   Schroeder,H. and Hauer,B.
  TITLE     METHOD FOR PRODUCING BIOTIN
  JOURNAL   Patent: WO 9905285-A 9 04-FEB-1999;
            SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES             Location/Qualifiers
```

```
     source          1. .3465
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /clone="PHS2BIOS1"
     CDS             272. .1477
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB77089.1"
                     /db_xref="GI:7240888"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT      804 a    963 c    934 g    764 t
ORIGIN

  Query Match           100.0%;  Score 1206;  DB 6;  Length 3465;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-301;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     272 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 331

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     332 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 391

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 451

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     452 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 511

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     512 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 571

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     572 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 631

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     632 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 691

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     692 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 751

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     752 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCCGGATCTGGCG 811

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     812 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 871

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     872 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 931

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     932 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 991

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     992 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1051

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1052 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1111

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1112 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1171

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1172 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1231

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1291

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1351

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1411

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1471

Qy    1201 GATTAA 1206
          ||||||
Db    1472 GATTAA 1477


RESULT 4
AX000474
LOCUS       AX000474               3794 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 5 from Patent WO9905285.
ACCESSION   AX000474
VERSION     AX000474.1  GI:7240883
KEYWORDS    .
SOURCE      unidentified.
  ORGANISM  unidentified
            unclassified.
REFERENCE   1  (bases 1 to 3794)
  AUTHORS   Schroeder,H. and Hauer,B.
  TITLE     METHOD FOR PRODUCING BIOTIN
  JOURNAL   Patent: WO 9905285-A 5 04-FEB-1999;
            SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES             Location/Qualifiers
     source          1. .3794
                     /organism="unidentified"
                     /db_xref="taxon:32644"
                     /clone="PHS1BIOS1"
     CDS             601. .1806
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAB77087.1"
                     /db_xref="GI:7240884"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT      910 a    1034 c    1020 g    830 t
ORIGIN

  Query Match           100.0%;  Score 1206;  DB 6;  Length 3794;
```

```
  Best Local Similarity   100.0%;  Pred. No. 1.1e-301;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 660

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 720

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 780

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 840

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 900

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 960

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 1020

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 1080

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 1140

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 1200

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 1260

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 1320

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1380

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1440

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1500

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1560

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1620

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1680

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1740

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1800

Qy    1201 GATTAA 1206
           ||||||
Db    1801 GATTAA 1806


RESULT 5
AX018970
LOCUS       AX018970               3794 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 11 from Patent WO9942591.
ACCESSION   AX018970
VERSION     AX018970.1  GI:10043065
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 3794)
  AUTHORS   Schroeder,H.
  TITLE     Method for producing biotin
  JOURNAL   Patent: WO 9942591-A 11 26-AUG-1999;
            SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES             Location/Qualifiers
     source          1. .3794
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             601. .1806
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC07719.1"
                     /db_xref="GI:10043066"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT      910 a   1034 c   1020 g    830 t
ORIGIN

  Query Match           100.0%;  Score 1206;  DB 6;  Length 3794;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-301;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 660

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 720

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 780

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 840

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
```

```
Db    841 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 900

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 960

Qy    361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 1020

Qy    421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 1080

Qy    481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 1140

Qy    541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 1200

Qy    601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 1260

Qy    661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 1320

Qy    721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1380

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1440

Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1500

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1560

Qy    961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1620

Qy   1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1680

Qy   1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1740

Qy   1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1800

Qy   1201 GATTAA 1206
          ||||||
Db   1801 GATTAA 1806


RESULT 6
AX018972
LOCUS       AX018972               4975 bp    DNA     linear   PAT 07-SEP-2000
DEFINITION  Sequence 13 from Patent WO9942591.
ACCESSION   AX018972
VERSION     AX018972.1  GI:10043067
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 4975)
  AUTHORS   Schroeder,H.
  TITLE     Method for producing biotin
  JOURNAL   Patent: WO 9942591-A 13 26-AUG-1999;
            SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES             Location/Qualifiers
     source          1. .4975
                     /organism="Escherichia coli"
                     /db_xref="taxon:562"
     CDS             530. .1684
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC07720.1"
                     /db_xref="GI:10043068"
                     /translation="MAKHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACET
                     YVKTGMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAIGKQSPD
                     INQGVDRADPLEQGAGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAEVRKNGTLPW
                     LRPDAKSQVTFQYDDGKIVGIDAVVLSTQHSEEIDQKSLQEAVMEEIIKPILPAEWLT
                     SATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTYGGMARHGGGAFSGKDPSKVDRSAA
                     YAARYVAKNIVAAGLADRCEIQVSYAIGVAEPTSIMVETFGTEKVPSEQLTLLVREFF
                     DLRPYGLIQMLDLLHPIYKETAAYGHFGREHFPWEKTDKAQLLRDAAGLK"
     CDS             1782. .2987
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAC07721.1"
                     /db_xref="GI:10043069"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLIARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT     1198 a   1350 c   1329 g   1098 t
ORIGIN

  Query Match           100.0%;  Score 1206;  DB 6;  Length 4975;
  Best Local Similarity  100.0%;  Pred. No. 1.1e-301;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1782 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCGCACTACAGGATGCGGGC 1841

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1842 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 1901

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1902 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 1961

Qy    181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1962 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 2021

Qy    241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2022 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 2081

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2082 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 2141
```

```
Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2142 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 2201

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2202 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 2261

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2262 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 2321

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2322 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 2381

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2382 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 2441

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2442 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 2501

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2502 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 2561

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2562 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 2621

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2622 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 2681

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2682 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 2741

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2742 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 2801

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2802 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 2861

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2862 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 2921

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2922 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 2981

Qy    1201 GATTAA 1206
           ||||||
Db    2982 GATTAA 2987
```

```
RESULT 7
AE000364
LOCUS       AE000364               12058 bp    DNA     linear   BCT 01-DEC-2000
DEFINITION  Escherichia coli K12 MG1655 section 254 of 400 of the complete
            genome.
ACCESSION   AE000364 U00096
VERSION     AE000364.1  GI:2367162
KEYWORDS    .
SOURCE      Escherichia coli K12.
  ORGANISM  Escherichia coli K12
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 12058)
  AUTHORS   Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V.,
            Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F.,
            Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J.,
            Mau,B. and Shao,Y.
  TITLE     The complete genome sequence of Escherichia coli K-12
  JOURNAL   Science 277 (5331), 1453-1474 (1997)
  MEDLINE   97426617
   PUBMED   9278503
REFERENCE   2  (bases 1 to 12058)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   3  (bases 1 to 12058)
  AUTHORS   Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
            University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
            Email: ecoli@genetics.wisc.edu Phone: 608-262-2534 Fax:
            608-263-7459
REFERENCE   4  (bases 1 to 12058)
  AUTHORS   Plunkett,G. III.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-OCT-1998) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     On Sep 9, 1997 this sequence version replaced gi:1789165.
            This sequence was determined by the E. coli Genome Project at the
            University of Wisconsin-Madison (Frederick R. Blattner, director).
            Supported by NIH grants HG00301 and HG01428 (from the Human Genome
            Project and NCHGR). The entire sequence was independently
            determined from E. coli K12 strain MG1655. Predicted open reading
            frames were determined using GeneMark software, kindly supplied by
            Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
            30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that
            have been correlated with genetic loci are being annotated with CG
            Site Nos., unique ID nos. for the genes in the E. coli Genetic
            Stock Center (CGSC) database at Yale University, kindly supplied by
            Mary Berlyn. A public version of the database is accessible
            (http://cgsc.biology.yale.edu). Annotation of the genome is an
            ongoing task whose goal is to make the genome sequence more useful
            by correlating it with other data. Comments to the authors are
            appreciated. Updated information will be available at the E. coli
            Genome Project's World Wide Web site
            (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
            its annotations are periodically updated; this is version M54. No
            sequence changes. Annotation updates: updated gene identifications
            and products; all new functional assignments courtesy of Monica
            Riley; added promoters, protein binding sites, and repeated
            sequences described in reference 1. The unique numeric identifiers
            beginning with a lowercase 'b' assigned to each gene (protein- or
            RNA-encoding) are now designated as gene synonyms instead of
            labels. This should allow them to be searched for in Entrez as gene
            names.
FEATURES             Location/Qualifiers
     source          1. .12058
                     /organism="Escherichia coli K12"
                     /strain="K12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:83333"
     protein_bind    45. .67
                     /note="central position to predicted promoter: -231.5"
                     /bound_moiety="ArgR predicted site"
     promoter        252. .282
                     /note="factor Sigma70; predicted +1 start at 2932242"
     gene            304. .1620
                     /gene="fucP"
                     /note="b2801"
     CDS             304. .1620
                     /gene="fucP"
                     /function="transport; Transport of small molecules:
```

```
                Carbohydrates, organic acids, alcohols"
                /note="o438; 100 pct identical to FUCP_ECOLI SW: P11551;
                CG Site no. 10875"
                /codon_start=1
                /transl_table=11
                /product="fucose permease"
                /protein_id="AAC75843.1"
                /db_xref="GI:1789166"
                /translation="MGNTSIQTQSYRAVDKDAGQSRSYIIPFALLCSLFFLWAVANNL
                NDILLPQFQQAFTLTNFQAGLIQSAFYFGYFIIPIPAGILMKKLSYKAGIITGLFLYA
                LGAALFWPAAEIMNYTLFLVGLFIIAAGLGCLETAANPFVTVLGPESSGHFRLNLAQT
                FNSFGAIIAVVFGQSLILSNVPHQSQDVLDKMSPEQLSAYKHSLVLSVQTPYMIIVAI
                VLLVALLIMLTKFPALQSDNHSDAKQGSFSASLSRLARIRHWRWAVLAQFCYVGAQTA
                CWSYLIRYAVEEIPGMTAGFAANYLTGTMVCFFIGRFTGTWLISRFAPHKVLAAYALI
                AMALCLISAFAQGHVGLIALTLCSAFMSIQYPTIFSLGIKNLGQDTKYGSSFIVMTII
                GGGIVTPVMGFVSDAAGNIPTAELIPALCFAVIFIFARFRSQTATN"
gene            1653. .3428
                /gene="fucI"
                /note="b2802"
CDS             1653. .3428
                /gene="fucI"
                /EC_number="5.3.1.-"
                /function="enzyme; Degradation of small molecules: Carbon
                compounds"
                /note="o591; 100 pct identical to FUCI_ECOLI SW: P11552;
                CG Site no. 10878"
                /codon_start=1
                /transl_table=11
                /product="L-fucose isomerase"
                /protein_id="AAC75844.1"
                /db_xref="GI:1789167"
                /translation="MKKISLPKIGIRPVIDGRRMGVRESLEEQTMNMAKATAALLTEK
                LRHACGAAVECVISDTCIAGMAEAAACEEKFSSQNVGLTITVTPCWCYGSETIDMDPT
                RPKAIWGFNGTERPGAVYLAAALAAHSQKGIPAFSIYGHDVQDADDTSIPADVEEKLL
                RFARAGLAVASMKGKSYLSLGGVSMGIAGSIVDHNFFESWLGMKVQAVDMTELRRRID
                QKIYDEAELEMALAWADKNFRYGEDENNKQYQRNAEQSRAVLRESLLMAMCIRDMMQG
                NSKLADIGRVEESLGYNAIAAGFQGQRHWTDQYPNGDTAEAILNSSFDWNGVREPFVV
                ATENDSLNGVAMLMGHQLTGTAQVFADVRTYWSPEAIERVTGHKLDGLAEHGIIHLIN
                SGSAALDGSCKQRDSEGNPTMKPHWEISQQEADACLAATEWCPAIHEYFRGGGYSSRF
                LTEGGVPFTMTRVNIIKGLGPVLQIAEGWSVELPKDVHDILNKRTNSTWPTTWFAPRL
                TGKGPFTDVYSVMANWGANHGVLTIGHVGADFITLASMLRIPVCMHNVEETKVYRPSA
                WAAHGMDIEGQDYRACQNYGPLYKR"
gene            3507. .4955
                /gene="fucK"
                /note="b2803"
CDS             3507. .4955
                /gene="fucK"
                /EC_number="2.7.1.51"
                /function="enzyme; Degradation of small molecules: Carbon
                compounds"
                /note="o482; 100 pct identical to FUCK_ECOLI SW: P11553;
                CG Site No. 19878"
                /codon_start=1
                /transl_table=11
                /product="L-fuculokinase"
                /protein_id="AAC75845.1"
                /db_xref="GI:1789168"
                /translation="MLSGYIAGAIMKQEVILVLDCGATNVRAIAVNRQGKIVARASTP
                NASDIAMENNTWHQWSLDAILQRFADCCRQINSELTECHIRGIAVTTFGVDGALVDKQ
                GNLLYPIISWKCPRTAAVMDNIERLISAQRLQAISGVGAFSFNTLYKLVWLKENHPQL
                LERAHAWLFISSLINHRLTGEFTTDITMAGTSQMLDIQQRDFSPQILQATGIPRRLFP
                RLVEAGEQIGTLQNSAAAMLGLPVGIPVISAGHDTQFALFGAGAEQNEPVLSSGTWEI
                LMVRSAQVDTSLLSQYAGSTCELDSQAGLYNPGMQWLASGVLEWVRKLFWTAETPWQM
                LIEEARLIAPGADGVKMQCDLLSCQNAGWQGVTLNTTRGHFYRAALEGLTAQLQRNLQ
                MLEKIGHFKASELLLVGGGSRNTLWNQIKANMLDIPVKVLDDAETTVAGAALFGWYGV
                GEFNSPEEARAQIHYQYRYFYPQTEPEFIEEV"
gene            4957. .5379
                /gene="fucU"
                /note="b2804"
CDS             4957. .5379
                /gene="fucU"
                /function="phenotype; Degradation of small molecules:
                Carbon compounds"
                /note="o140; 100 pct identical to FUCU_ECOLI SW: P11555
                but35 C-terminal aa are dissimilar to reported 28 aa"
                /codon_start=1
                /transl_table=11
                /product="protein of fucose operon"
                /protein_id="AAC75846.1"
                /db_xref="GI:1789169"
                /translation="MLKTISPLISPELLKVLAEMGHGDEIIFSDAHFPAHSMGPQVIR
                ADGLLVSDLLQAIIPLFELDSYAPPLVMMAAVEGDTLDPEVERRYRNALSLQAPCPDI
                IRINRFAFYERAQKAFAIVITGERAKYGNILLKKGVTP"
gene            5437. .6168
                /gene="fucR"
                /note="b2805"
CDS             5437. .6168
                /gene="fucR"
                /function="regulator; Degradation of small molecules:
                Carbon compounds"
                /note="o243; 100 pct identical to FUCR_ECOLI SW: P11554;
                CG Site No. 10884"
                /codon_start=1
                /transl_table=11
                /product="positive regulator of the fuc operon"
                /protein_id="AAC75847.1"
                /db_xref="GI:1789170"
                /translation="MKAARQQAIVDLLLNHTSLTTEALSEQLKVSKETIRRDLNELQT
                QGKILRNHGRAKYIHRQNQDSGDPFHIRLKSHYAHKADIAREALAWIEEGMVIALDAS
                STCWYLARQLPDINIQVFTNSHPICHELGKRERIQLISSGGTLERKYGCYVNPSLISQ
                LKSLEIDLFIFSCEGIDSSGALWDSNAINADYKSMLLKRAAQSLLLIDKSKFNRSGEA
                RIGHLDEVTHIISDERQVATSLVTA"
gene            complement(6212. .7312)
                /gene="ygdE"
                /note="b2806"
CDS             complement(6212. .7312)
                /gene="ygdE"
                /function="orf; Unknown"
                /note="f366; 100 pct identical to YGDE_ECOLI SW: P32066;
                alternate name Orf2 of X73413 and U01030"
                /codon_start=1
                /transl_table=11
                /product="orf, hypothetical protein"
                /protein_id="AAC75848.1"
                /db_xref="GI:1789171"
                /translation="MNKVVLLCRPGFEKECAAEITDKAGQREIFGFARVKENAGYVIY
                ECYQPDDGDKLIRELPFSSLIFARQWFVVGELLQHLPPEDRITPIVGMLQGVVEKGGE
```

```
Query Match           100.0%;  Score 1206;  DB 1;  Length 12058;
Best Local Similarity 100.0%;  Pred. No. 1.1e-301;
Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9406 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 9465

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9466 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 9525

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9526 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 9585

Qy    181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9586 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 9645

Qy    241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9646 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 9705

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9706 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 9765
```

```
Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9766 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 9825

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9826 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 9885

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9886 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 9945

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9946 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 10005

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10006 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 10065

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10066 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 10125

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10126 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 10185

Qy     781 ACCACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10186 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 10245

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10246 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 10305

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10306 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 10365

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10366 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 10425

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10426 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 10485

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10486 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 10545

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   10546 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 10605

Qy    1201 GATTAA 1206
           ||||||
Db   10606 GATTAA 10611
```

RESULT 8
ECU29581

```
LOCUS       ECU29581              71128 bp    DNA     linear   BCT 14-JAN-1997
DEFINITION  Escherichia coli K-12 genome; approximately 63 to 64 minutes.
ACCESSION   U29581
VERSION     U29581.1  GI:1255723
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 71128)
  AUTHORS   Reizer,J., Reizer,A., Merrick,M.J., Plunkett,G. III, Rose,D.J. and
            Saier,M.H. Jr.
  TITLE     Novel phosphotransferase-encoding genes revealed by analysis of the
            Escherichia coli genome: a chimeric gene encoding an Enzyme I
            homologue that possesses a putative sensory transduction domain
  JOURNAL   Gene 181 (1-2), 103-108 (1996)
  MEDLINE   97128775
   PUBMED   8973315
REFERENCE   2  (bases 1 to 71128)
  AUTHORS   Plunkett,G.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUN-1995) Guy Plunkett, University of Wisconsin,
            Laboratory of Genetics, 445 Henry Mall, Madison, WI 53706, USA
COMMENT     On Apr 5, 1996 this sequence version replaced gi:882681.
            This sequence was determined as part of the E. coli Genome Project
            (Frederick R. Blattner, director) at the University of
            Wisconsin-Madison. Supported by award HG00301 from the NIH Human
            Genome Project. The entire sequence was independently determined
            from E. coli MG1655; overlaps with other sequence determinations
            are annotated. This entry should be considered provisional.
            Reference [1] describes a correction and update to the origional
            entry's sequence and annotations.
FEATURES             Location/Qualifiers
     source          1. .71128
                     /organism="Escherichia coli"
                     /strain="K-12"
                     /sub_strain="MG1655"
                     /db_xref="taxon:562"
                     /map="approximately 63 to 64 minutes"
                     /note="This sequence comprises the following lambda
                     clones: DD628(EC18-104), DD632(EC30K458-4), DD635(EC18-
                     251), DD639(EC30MM20), and DD641(EC17TWR8); M13Janus
                     vector was used for subcloning"
     CDS             complement(<1. .940)
                     /note="ORF_f313p"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB40437.1"
                     /db_xref="GI:882682"
                     /translation="MSSQFTTPVVTEMQVIPVAGHDSMLMNLSGAHAPFFTRNIVIIK
                     DNSGHTGVGEIPGGEKIRKTLEDAIPLVVGKTLGEYKNVLTLVRNTFADRDAGGRGLQ
                     TFDLRTTIHVVTGIEAAMLDLLGQHLGVNVASLLGDGQQRSEVEMLGYLFFVGNRKAT
                     PLPYQSQPDDSCDWYRLRHEEAMTPDAVVRLAEAAYEKYGFNDFKLKGGVLAGEEEAE
                     SIVALAQRFPQARITLDPNGAWSLNEAIKIGKYLKGSLAYAEDPCGAEQGFSGREVMA
                     EFRRATGLPTATNMIATDWRQMGHTLSLQSVDIRWRI"
     CDS             complement(961. .2301)
                     /note="ORF_f446"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB40438.1"
                     /db_xref="GI:882683"
                     /translation="MATQSSPVITDMKVIPVAGHDSMLLNIGGAHNAYFTRNIVVLTD
                     NAGHTGIGEAPGGDVIYQTLVDAIPMVLGQEVARLNKVVQQVHKGNQAADFDTFGKGA
                     WTFELRVNAVAALEAALLDLLGKALNVPVCELLGPGKQREAITVLGYLFYIGDRTKTD
                     LPYVENTPGNHEWYQLRHQKAMNSEAVVRLAEASQDRYGFKDFKLKGGVLPGEQEIDT
                     VRALKKRFPDARITVDPNGAWLLDEAISLCKGLNDVLTYAEDPCGAEQGFSGREVMAE
                     FRRATGLPVATNMIATNWREMGHAVMLNAVDIPLADPHFWTLSGAVRVAQLCDDWGLT
                     WGCHSNNHFDISLAMFTHVGAAAPGNPTAIDTHWIWQEGDCRLTQNPLEIKNGKIAVP
                     DAPGLGVELDWEQVQKAHEAYKRLPGGARNDAGPMQYLIPGWTFDRKRPVFGRH"
     CDS             complement(2303. .3655)
                     /note="ORF_f450"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="AAB40439.1"
                     /db_xref="GI:882684"
                     /translation="MSSLSQAASSVEKRTNARYWIVVMLFIVTSFNYGDRATLSIAGS
                     EMAKDIGLDPVGMGYVFSAFSWAYVIGQIPGGWLLDRFGSKRVYFWSIFIWSMFTLLQ
                     GFVDIFSGFGIIVALFTLRFLVGLAEAPSFPGNSRIVAAWFPAQERGTAVSIFNSAQY
                     FATVIFAPIMGWLTHEVGWSHVFFFMGGLGIVISFIWLKVIHEPNQHPGVNKKELEYI
                     AAGGALINMDQQNTKVKVPFSVKWGQIKQLLGSRMMIGVYIGQYCINALTYFFITWFP
                     VYLVQARGMSILKAGFVASVPAVCGFIGGVLGGIISDWLMRRTGSLNIARKTPIVMGM
                     LLSMVMVFCNYVNVEWMIIGFMALAFFGKGIGALGWAVMADTAPKEISGLSGGLFNMF
                     GNISGIVTPIAIGYIVGTTGSFNGALIYVGVHALIAVLSYLVLVGDIKRIELKPVAGQ
```

```
                "
CDS             complement(4090. .4539)
                /note="corresponds to hypothetical protein from E.
                carotovora PIR Accession Number S45107; ORF_f150"
                /codon_start=1
                /transl_table=11
                /protein_id="AAB40440.1"
                /db_xref="GI:882685"
                /translation="MAEIGIFVGTMYGNSLLVAEEAEAILTAQGHKATVFEDPELSDW
                LPYQDKYVLVVTSTTGQGDLPDSIVPLFQGIKDSLGFQPNLRYGVIALGDSSYVNFCN
                GGKQFDALLQEQSAQRVGEMLLIDASENPEPETESNPWVEQWGTLLS"
CDS             complement(4557. .5339)
                /note="corresponds to hypothetical protein from E.
                carotovora PIR Accession Number S45108; 2 ORFs in E. coli;
                ORF_f260"
                /codon_start=1
                /transl_table=11
                /protein_id="AAB40441.1"
                /db_xref="GI:882686"
                /translation="MLEILYQDEWLVAVNKPSGWLVHRSWLDRDEKVVVMQTVRDQIG
                QHVFTAHRLDRPTSGVLLMGLSSEAGRLLAQQFEQHQIQKRYHAIVRGWLMEEAVLDY
                PLVEELDKIADKFAREDKGPQPAVTHYRGLATVEMPVATGRYPTTRYGLVELEPKTGR
                KHQLRRHLAHLRHPIIGDSKHGDLRQNRSGAEHFGLQRLMLHASQLSLTHPFTGEPLT
                IHAGLDDTWMQALSQFGWRGLLPENERVEFSAPSGQDGEISS"
CDS             complement(5339. .5668)
                /note="corresponds to hypothetical protein from E.
                carotovora PIR Accession Number S45108; 2 ORFs in E. coli;
                ORF_f109"
                /codon_start=1
                /transl_table=11
                /protein_id="AAB40442.1"
                /db_xref="GI:882687"
                /translation="MTTHDRVRLQLQALEALLREHQHWRNDEPQPHQFNSTQPFFMDT
                MEPLEWLQWVLIPRMHDLLDNKQPLPGAFAVAPYYEMALATDHPQRALILAELEKLDA
                LFADDAS"
CDS             complement(6290. .6835)
                /note="GTG start"
                /codon_start=1
                /transl_table=11
                /protein_id="AAB40443.1"
                /db_xref="GI:882688"
                /translation="MDDLTAQALKDFTARYCDAWHEEHKSWPLSEELYGVPSPCIIST
                TEDAVYWQPQPFTGEQNVNAVERAFDIVIQPTIHTFYTTQFAGDMHAQFGDIKLTLLQ
                TWSEDDFRRVQENLIGHLVTQKRLKLPPTLFIATLEEELEVISVCNLSGEVCKETLGT
                RKRTHLASNLAEFLNQLKPLL"
CDS             6903. .7751
                /note="ORF_o282"
                /codon_start=1
                /transl_table=11
                /protein_id="AAB40444.1"
                /db_xref="GI:882689"
                /translation="MSSYANHQALAGLTLGKSTDYRDTYDASLLQGVPRSLNRDPLGL
                KADNLPFHGTDIWTLYELSWLNAKGLPQVAVGHVELDYTSVNLIESKSFKLYLNSFNQ
                TRFNNWDEVRQTLERDLSTCAQGKISVALYRLDELEGQPIGHFNGTCIDDQDITIDNY
                EFTTDYLENATCGEKVVEETLVSHLLKSNCLITHQPDWGSLQIQYRGRQIDREKLLRY
                LVSFRHHNEFHEQCVERIFNDLLRFCQPEKLSVYARYTRRGGLDINPWRSNSDFVPST
                TRLVRQ"
CDS             7863. .9227
                /note="TTG start; alternate gene name ygdH; ORF_o455"
                /codon_start=1
                /transl_table=11
                /protein_id="AAB40445.1"
                /db_xref="GI:882690"
                /translation="MITHISPLGSMDMLSQLEVDMLKRTASSDLYQLFRNCSLAVLNS
                GSLTDNSKELLSRFENFDINVLRRERGVKLELINPPEEAFVDGRIIRALQANLFAVLR
                DILFVYGQIHNTVRFPNLNLDNSVHITNLVFSILRNARALHVGEAPNMVVCWGGHSIN
                ENEYLYARRVGNQLGLRELNICTGCGPGAMEAPMKGAAVGHAQQRYKDSRFIGMTEPS
                IIAAEPPNPLVNELIIMPDIEKRLEAFVRIAHGIIIFPGGVGTAEELLYLLGILMNPA
                NKDQVLPLILTGPKESADYFRVLDEFVVHTLGENARRHYRIIIDDAAEVARQMKKSMP
                LVKENRRDTGDAYSFNWSMRIAPDLQMPFEPSHENMANLKLYPDQPVEVLAADLRRAF
                SGIVAGNVKEVGIRAIEEFGPYKINGDKEIMRRMDDLLQGFVAQHRMKLPGSAYIPCY
                EICT"
misc_feature    8982. .11136
                /note="similar to GenBank Accession Number U01233
                (ECU01233) 1. .2155"
gene            9784. .11073
                /gene="sdaC"
CDS             9784. .11073
                /gene="sdaC"
                /note="CG Site no. 33329; ORF_o429"
                /codon_start=1
                /transl_table=11
                /product="putative serine transporter"
                /protein_id="AAB40446.1"
                /db_xref="GI:882691"
                /translation="METTQTSTIASKDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPIN
                AGVGGMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLIT
                LLYFFAIYPILLVYSVAITNTVESFMSHQLGMTPPPRAILSLILIVGMMTIVRFGEQM
                IVKAMSILVFPFVGVLMLLALYLIPQWNGAALETLSLDTASATGNGLWMTLWLAIPVM
                VFSFNHSPIISSFAVAKREEYGDMAEQKCSKILAFAHIMMVLTVMFFVFSCVLSLTPA
                DLAAAKEQNISILSYLANHFNAPVIAWMAPIIAIIAITKSFLGHYLGAREGFNGMVIK
                SLRGKGKSIEINKLNRITALFMLVTTWIVATLNPSILGMIETLGGPIIAMILFLMPMY
                AIQKVPAMRKYSGHISNVFVVVMGLIAISAIFYSLFS"
misc_feature    10731. .13515
                /note="similar to GenBank Accession Number L07763
                (ECOSDAB) 1. .2785"
gene            11131. .12498
                /gene="sdaB"
CDS             11131. .12498
                /gene="sdaB"
                /EC_number="4.2.1.13"
                /note="CG Site no. 33324; ORF_o455"
                /codon_start=1
                /transl_table=11
                /product="L-serine dehydratase 2 (L-serine deaminase 2)"
                /protein_id="AAB40447.1"
                /db_xref="GI:882692"
                /translation="MISVFDIFKIGIGPSSSHTVGPMKAGKQFTDDLIARNLLKDVTR
                VVVDVYGSLSLTGKGHHTDIAIIMGLAGNLPDTVDIDSIPSFIQDVNTHGRLMLANGQ
                HEVEFPVDQCMNFHADNLSLHENGMRITALAGDKVVYSQTYYSIGGGFIVDEEHFGQQ
                DSAPVEVPYPYSSAADLQKHCQETGLSLSGLMMKNELALHSKEELEQHLANVWEVMRG
                GIERGISTEGVLPGKLRVPRRAAALRRMLVSQDKTTTDPMAVVDWINMFALAVNEENA
                AGGRVVTAPTNGACGIIPAVLAYYDKFIREVNANSLARYLLVASAIGSLYKMNASISG
                AEVGCQGEVGVGSMAAAGLAELLGASPAQVCIAAEIAMEHNLGLTCDPVAGQVQVPC
                IERNAIAAVKAVNAARMALRRTSEPRVCLDKVIETMYETGKDMNAKYRETSRGGLAMK
                IVACD"
CDS             12520. .13365

  Query Match          100.0%;  Score 1206;  DB 1;  Length 71128;
  Best Local Similarity  100.0%;  Pred. No. 1.2e-301;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24892 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 24951

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   24952 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 25011

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25012 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 25071

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25072 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 25131

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25132 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 25191

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25192 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 25251
```

```
Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25252 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 25311

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25312 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 25371

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25372 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 25431

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25432 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 25491

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25492 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 25551

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25552 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 25611

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25612 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 25671

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25672 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 25731

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25732 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 25791

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25792 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 25851

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25852 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 25911

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25912 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 25971

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   25972 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 26031

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   26032 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 26091

Qy    1201 GATTAA 1206
           ||||||
Db   26092 GATTAA 26097
```

```
RESULT 9
AX000470
LOCUS       AX000470               1216 bp    DNA     linear   PAT 10-MAR-2000
DEFINITION  Sequence 1 from Patent WO9905285.
ACCESSION   AX000470
VERSION     AX000470.1  GI:7240879
KEYWORDS    .
SOURCE      Escherichia coli.
  ORGANISM  Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 1216)
  AUTHORS   Schroeder,H. and Hauer,B.
  TITLE     METHOD FOR PRODUCING BIOTIN
  JOURNAL   Patent: WO 9905285-A 1 04-FEB-1999;
            SCHROEDER HARTWIG (DE); BASF AG (DE)
FEATURES             Location/Qualifiers
     source          1. .1216
                     /organism="Escherichia coli"
                     /strain="W3110"
                     /db_xref="taxon:562"
     5'UTR           1. .11
     CDS             12. .>1216
                     /note="unnamed protein product"
                     /codon_start=1
                     /transl_table=11
                     /protein_id="CAB77085.1"
                     /db_xref="GI:7240880"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATQQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGMVVMVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSAALEWLADYDINQAESWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
BASE COUNT      258 a    318 c    363 g    277 t
ORIGIN
```

```
  Query Match             99.9%;  Score 1205;  DB 6;  Length 1216;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-301;
  Matches 1205;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 71

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 131

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 191

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     192 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 251

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     252 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 311

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     312 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 371

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     372 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 431

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 491

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     492 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 551

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     552 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 611

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     612 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 671

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     672 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 731

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     732 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 791

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     792 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 851

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     852 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 911

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     912 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 971

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     972 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1031

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1032 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1091

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1092 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1151

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCGCTGGAATTATTGGTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1152 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCGCTGGAATTATTGGTG 1211

Qy    1201 GATTA 1205
           |||||
Db    1212 GATTA 1216
```

```
RESULT 10
AP002563
LOCUS       AP002563               266658 bp    DNA     linear   BCT 07-MAR-2001
DEFINITION  Escherichia coli O157:H7 DNA, complete genome, section 14/20.
ACCESSION   AP002563 BA000007
VERSION     AP002563.1  GI:13363121
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 (strain:O157:H7, sub_strain:RIMD 0509952)
            DNA.
  ORGANISM  Escherichia coli O157:H7
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (sites)
  AUTHORS   Makino,K., Yokoyama,K., Kubota,Y., Yutsudo,C.H., Kimura,S.,
            Kurokawa,K., Ishii,K., Hattori,M., Tatsuno,I., Abe,H., Iida,T.,
            Yamamoto,K., Ohnishi,M., Hayashi,T., Yasunaga,T., Honda,T.,
            Sasakawa,C. and Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT2-Sakai carrying the
            verotoxin 2 genes of the enterohemorrhagic Escherichia coli O157:H7
            derived from the Sakai outbreak
  JOURNAL   Genes Genet. Syst. 74 (5), 227-239 (1999)
  MEDLINE   20198780
REFERENCE   2  (sites)
  AUTHORS   Ohnishi,M., Murata,T., Nakayama,K., Kuhara,S., Hattori,M.,
            Kurokawa,K., Yasunaga,T., Yokoyama,K., Makino,K., Shinagawa,H. and
            Hayashi,T.
  TITLE     Comparative analysis of the whole set of rRNA operons between an
            enterohemorrhagic Escherichia coli O157:H7 Sakai strain and an
            Escherichia coli K-12 strain MG1655
  JOURNAL   Syst. Appl. Microbiol. 23 (3), 315-324 (2000)
  MEDLINE   20557356
REFERENCE   3  (sites)
  AUTHORS   Yokoyama,K., Makino,K., Kubota,Y., Watanabe,M., Kimura,S.,
            Yutsudo,C.H., Kurokawa,K., Ishii,K., Hattori,M., Abe,H., Iida,T.,
            Yamamoto,K., Hayashi,T., Yasunaga,T., Honda,T., Sasakawa,C. and
            Shinagawa,H.
  TITLE     Complete nucleotide sequence of the prophage VT1-Sakai carrying the
            Shiga toxin 1 genes of the enterohemorrhagic Escherichia coli
            O157:H7 strain derived from the Sakai outbreak
  JOURNAL   Gene 258 (1-2), 127-139 (2000)
  MEDLINE   20564182
REFERENCE   4  (sites)
  AUTHORS   Hayashi,T., Makino,K., Ohnishi,M., Kurokawa,K., Ishii,K.,
            Yokoyama,K., Han,C.-G., Ohtsubo,E., Nakayama,K., Murata,T.,
            Tanaka,M., Tobe,T., Iida,T., Takami,H., Honda,T., Sasakawa,C.,
            Ogasawara,N., Yasunaga,T., Kuhara,S., Shiba,T., Hattori,M. and
            Shinagawa,H.
  TITLE     Complete genome sequence of enterohemorrhagic Escherichia coli
            O157:H7 and genomic comparison with a laboratory strain K-12
  JOURNAL   DNA Res. 8 (1), 11-22 (2001)
  MEDLINE   21156231
REFERENCE   5  (bases 1 to 266658)
  AUTHORS   Ohnishi,M., Kurokawa,K., Makino,K., Yasunaga,T., Shinagawa,H. and
            Hayashi,T.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JUN-2000) Ken Kurokawa, Osaka University, Genome
            Information Research Center; 3-1, Yamadaoka, Suita, Osaka 565-0871,
            Japan (E-mail:ken@gen-info.osaka-u.ac.jp,
            URL:http://www.gen-info.osaka-u.ac.jp/, Tel:81-6-6879-8365,
            Fax:81-6-6879-2047)
COMMENT     genome project.
FEATURES             Location/Qualifiers
     source          1. .266658
                     /organism="Escherichia coli O157:H7"
                     /strain="O157:H7"
                     /sub_strain="RIMD 0509952"
                     /db_xref="taxon:83334"
     gene            complement(244. .693)
                     /gene="ECs3650"
     CDS             complement(244. .693)
                     /gene="ECs3650"
                     /note="similar to B2790_ECOLI gi|1789154 percent identity
                     100 in 149 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37073.1"
                     /db_xref="GI:13363122"
                     /translation="MAEIGIFVGTMYGNSLLVAEEAEAYLTAQGHKATVFEDPELSDW
                     LPYQDKYVLVVTSTTGQGDLPDSIVPLFQGIKDSLGFQPNLRYGVIALGDSSYVNFCN
                     GGKQFDALLQEQSAQRVGEMLLIDASENPEPETESNPWVEQWGTLLS"
     gene            complement(711. .1493)
                     /gene="ECs3651"
     CDS             complement(711. .1493)
                     /gene="ECs3651"
                     /note="similar to YQCB_ECOLI gi|1789155 percent identity
                     99 in 260 aa (Conserved in E.coli K-12)"
                     /codon_start=1
                     /evidence=not_experimental
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="BAB37074.1"
                     /db_xref="GI:13363123"
                     /translation="MLEILYQDEWLVAVNKPSGWLVHRSWLDRDEKVVVMQTVRDQIG
                     QHVFTAHRLDRPTSGVLLMGLSSEAGRLLAQQFEQHQIQKRYHAIVRGWLMEEAVLDY
                     PLVEELDKIADKFAREDKGPQPAVTHYRGLATVEMPVATGRYPTTRYGLVELEPKTGR
                     KHQLRRHLAHLRHPIIGDSKHGDLRQNRSGAEHFGLQRLMLHASQLSLTHPFTGEPLT
                     IHAGLDDTWMQALSQFGWRGLLPENERVEFSAPSGQDGERSS"
     gene            complement(1493. .1822)
                     /gene="ECs3652"
     CDS             complement(1493. .1822)
                     /gene="ECs3652"
```

```
                /note="similar to B2792_ECOLI gi|1789156 percent identity
                98 in 109 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB37075.1"
                /db_xref="GI:13363124"
                /translation="MTTHDRVRLQLQALEALLREHQHWRNDEPQPHQFNSTQPFFMDT
                MEPLEWLQWVLIPRMHDLLNNNQPLPGAFAVAPYYEMALATDHPQRALILAELEKLDA
                LFADDAS"
gene            complement(1865. .2233)
                /gene="csrB"
misc_RNA        complement(1865. .2233)
                /gene="csrB"
                /product="CsrB RNA"
                /note="inhibitor of CsrA"
gene            complement(2444. .2989)
                /gene="ECs3653"
CDS             complement(2444. .2989)
                /gene="ECs3653"
                /note="similar to SYD_ECOLI gi|1789157 percent identity 98
                in 181 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="Syd protein"
                /protein_id="BAB37076.1"
                /db_xref="GI:13363125"
                /translation="MDDLTAQALKDFTARYCDAWHEEHKSWPLSEELYGVPSPCIIST
                TEDAVYWQPQPFTGEQNVNAVERAFDIVIQPTIHTFYTTQFAGDMHAQFGDIKLTLLQ
                TWSEDDFRRVQENLIGHLVTQKRLKLPPTLFIATLEEELEVISVCNLSGEVCKETLGT
                RKRTHLAPNLAEFLNQLKPLL"
gene            3057. .3905
                /gene="ECs3654"
CDS             3057. .3905
                /gene="ECs3654"
                /note="similar to YQCD_ECOLI gi|1789158 percent identity
                98 in 282 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB37077.1"
                /db_xref="GI:13363126"
                /translation="MSSYANHQALAGLTLGKSTDYRDTYDASLLQGVPRSLNRDPLGL
                KADNLPFQGTDIWTLYELSWLNAKGLPQVAVGHVELDYTSVNLIESKSFKLYLNSFNQ
                TRFNNWDEVRQTLERDLSTCAQGEVSVALYRLDELEGQPIGHFNGTCIDDQDITIDNY
                EFTTDYLENATSGEKVVEETLVSHLLKSNCLITHQPDWGSIQIQYRGRQIDREKLLRY
                LVSFRHHNEFHEQCVERIFNDLLRFCQPEKLSVYARYTRRGGLDINPWRSNSDFVPST
                TRLVRQ"
gene            4017. .5381
                /gene="ECs3655"
CDS             4017. .5381
                /gene="ECs3655"
                /note="similar to YGDH_ECOLI gi|1789159 percent identity
                99 in 454 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="hypothetical protein"
                /protein_id="BAB37078.1"
                /db_xref="GI:13363127"
                /translation="MITHISPLGSMDMLSQLEVDMLKRTASSDLYQLFRNCSLAVLNS
                GSLTDNSKELLSRFENFDINVLRRERGVKLELINPPEEAFVDGRIIRALQANLFAVLR
                DILFVYGQIHNTVRFPNLNLDNSVHITNLVFSILRNARALHVGEAPNMVVCWGGHSIN
                ENEYLYARRVGNQLGLRELNICTGCGPGAMEAPMKGAAVGHAQQRYKDSRFIGMTEPS
                IIAAEPPNPLVNELIIMPDIEKRLEAFVRIAHGIIIFPGGVGTAEELLYLLGILMNPA
                NKDQVLPLILTGPKESADYFRVLDEFVVHTLGENARRHYRIIIDDAAEVARQMKKSMP
                LVKENRRDTGDAYSFNWSMRIAPDLQMPFEPSHENMANLKLYPDQPVEVLAADLRRAF
                SGIVAGNVKEVGIRAIEEFGPYKINGDKEIMRRMDDLLQGFVAQHRMKLPGSAYIPCY
                EICT"
gene            5938. .7227
                /gene="ECs3656"
CDS             5938. .7227
                /gene="ECs3656"
                /note="similar to SDAC_ECOLI gi|1789160 percent identity
                100 in 429 aa (Conserved in E.coli K-12)"
                /codon_start=1
                /evidence=not_experimental
                /transl_table=11
                /product="probable serine transporter"
                /protein_id="BAB37079.1"
                /db_xref="GI:13363128"
                /translation="METTQTSTIASKDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPIN
                AGVGGMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLIT
                LLYFFAIYPILLVYSVAITNTVESFMSHQLGMTPPPRAILSLILIVGMMTIVRFGEQM
                IVKAMSILVFPFVGVLMLLALYLIPQWNGAALETLSLDTASATGNGLWMTLWLAIPVM
                VFSFNHSPIISSFAVAKREEYGDMAEQKCSKILAFAHIMMVLTVMFFVFSCVLSLTPA
                DLAAAKEQNISILSYLANHFNAPVIAWMAPIIAIIAITKSFLGHYLGAREGFNGMVIK
                SLRGKGKSIEINKLNRITALFMLVTTWIVATLNPSILGMIETLGGPIIAMILFLMPMY
                AIQKVPAMRKYSGHISNVFVVVMGLIAISAIFYSLFS"
gene            7285. .8652
                /gene="ECs3657"
CDS             7285. .8652
                /gene="ECs3657"
                /note="similar to SDAA_ECOLI gi|1788116 percent identity
                77 in 455 aa, similar to SDAB_ECOLI gi|1789161 percent
                identity 99 in 455 aa (Conserved in E.coli K-12)"

  Query Match             96.8%;  Score 1167.6;  DB 1;  Length 266658;
  Best Local Similarity   98.0%;  Pred. No. 1.2e-291;
  Matches 1182;  Conservative    0;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db   21046 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCTCTACAGGATGCGGGC 21105

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   21106 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCGA 21165

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21166 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 21225

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21226 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 21285

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21286 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 21345

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21346 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 21405

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db   21406 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTCAAATTG 21465

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          |||||||||||||||||||||| |||||||||||| |||| ||||||||||||||||||||
Db   21466 CCGCTTAATGCGCAGCGACTGCCAGATGTCGATTTGCTGCCGGAACTGATTACTCCCCGT 21525

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          |||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   21526 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTAACTGGCGGTTGCCCGGATCTGGCG 21585

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          |||||||||||||||||||||||||||||||| |||||||| |||||||||||||||||||
Db   21586 CGAGCGATTACCTTTGCTCATTCAGCCGGGATAGTGGTGATAGTTGATGGTGCTCAGGGG 21645

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          || |||||||||||||||||||||||||||||||| |||||||||||| || |||
```

```
Db  21646 GCGGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATCGATTTCTATGCTTTCTCTGGT 21705

Qy    661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          |||||||||||||||||||| |||||||||||||||||||| ||||||||||||||||||
Db  21706 CACAAACTGTATGGCCCGACGGGTATCGGCGTGCTGTATGGCAAATCAGAACTGCTGGAA 21765

Qy    721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21766 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 21825

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db  21826 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCAGGAACGCCAAATGTGGCTGGTGTCATA 21885

Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db  21886 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAACTGG 21945

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          |||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db  21946 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCAGGCTTTCGTTCA 22005

Qy    961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db  22006 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCACAGCGAT 22065

Qy   1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          |||||||||||||||||||||||||||||| |||||||||||| ||||||||||||||
Db  22066 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCTTGCGGGCCGGGCAACATTGCGCTCAG 22125

Qy   1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db  22126 CCGCTACTGGCAGAATTAGGCGTGACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 22185

Qy   1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  22186 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 22245

Qy   1201 GATTAA 1206
          ||||||
Db  22246 GATTAA 22251


RESULT 11
AE005509
LOCUS       AE005509               6488 bp    DNA     linear   BCT 21-MAR-2001
DEFINITION  Escherichia coli O157:H7 EDL933 genome, contig 3 of 3, section 128
            of 290.
ACCESSION   AE005509 AE005174
VERSION     AE005509.1  GI:12517290
KEYWORDS    .
SOURCE      Escherichia coli O157:H7 EDL933.
  ORGANISM  Escherichia coli O157:H7 EDL933
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE   1  (bases 1 to 6488)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Genome sequence of enterohaemorrhagic Escherichia coli O157:H7
  JOURNAL   Nature 409 (6819), 529-533 (2001)
  MEDLINE   21074935
   PUBMED   11206551
REFERENCE   2  (bases 1 to 6488)
  AUTHORS   Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
            Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
            Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
            Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
            Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
            Welch,R.A. and Blattner,F.R.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-OCT-2000) Laboratory of Genetics, University of
            Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .6488
                     /organism="Escherichia coli O157:H7 EDL933"
                     /strain="EDL933"
                     /serotype="O157:H7"
                     /db_xref="taxon:155864"
                     /note="enterohemorrhagic"
     gene            complement(206. .436)
                     /gene="Z4126"
     CDS             complement(206. .436)
                     /gene="Z4126"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 76 of 76 are 100.00 pct identical to
                     residues 1 to 76 of 76 from Escherichia coli K-12 Strain
                     MG1655: B2809"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG57923.1"
                     /db_xref="GI:12517291"
                     /translation="MMKKTAAIISACMLTXXLSACSGSNYVMHTNDGRTIVSDGKPQT
                     DNDTGMISYKDANGNKQQINRXDVKEMVELDQ"
     gene            625. .1830
                     /gene="Z4127"
     CDS             625. .1830
                     /gene="Z4127"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 401 of 401 are 99.00 pct identical to
                     residues 1 to 401 of 401 from Escherichia coli K-12 Strain
                     MG1655: B2810"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG57924.1"
                     /db_xref="GI:12517292"
                     /translation="MNVFNPAQFRAQFPALQDAGVYLDSAATALKPEAVVEATRQFYS
                     LSAGNVHRSQFAEAQRLTARYEAAREKVAQLLNAPDDKTIVWTRGTTESINMVAQCYA
                     RPRLQPGDEIIVSVAEHHANLVPWLMVAQQTGAKVVKLPLNAQRLPDVDLLPELITPR
                     SRILALGQMSNVTGGCPDLARAITFAHSAGIVVIVDGAQGAVHFPADVQQLDIDFYAF
                     SGHKLYGPTGIGVLYGKSELLEAMSPWLGGGKMVHEVSFDGFTTQSAPWKLEAGTPNV
                     AGVIGLSXALEWLADYDINQAENWSRSLATLAEDALAKRPGFRSFRCQDSSLLAFDFA
                     GVHHSDMVTLLAEYGIALRAGQHCAQPLLAELGVTGTLRASFAPYNTKSDVDALVNAV
                     DRALELLVD"
     gene            1830. .2273
                     /gene="ygdK"
                     /note="Z4128"
     CDS             1830. .2273
                     /gene="ygdK"
                     /function="orf; Unknown function"
                     /note="Residues 1 to 147 of 147 are 100.00 pct identical
                     to residues 1 to 147 of 147 from Escherichia coli K-12
                     Strain MG1655: B2811"
                     /codon_start=1
                     /transl_table=11
                     /product="orf, hypothetical protein"
                     /protein_id="AAG57925.1"
                     /db_xref="GI:12517293"
                     /translation="MTNPQFAGHPFGTTVTAETLRNTFAPLTQWEDKYRQLIMLGKQL
                     PALPDELKAQAKEIAGCENRVWLGYTVAENGKMHFFGDSEGRIVRGLLAVLLTAVEGK
                     TAAELQAQSPLALFDELGLRAQLSASRSQGLNALSEAIIAATKQV"
     gene            complement(2324. .3130)
                     /gene="ygdL"
                     /note="Z4129"
     CDS             complement(2324. .3130)
                     /gene="ygdL"
                     /function="putative enzyme; Not classified"
                     /note="Residues 1 to 268 of 268 are 98.88 pct identical to
                     residues 1 to 268 of 268 from Escherichia coli K-12 Strain
                     MG1655: B2812"
                     /codon_start=1
```

```
                /transl_table=11
                /product="putative enzyme"
                /protein_id="AAG57926.1"
                /db_xref="GI:12517294"
                /translation="MSVVISDAWRQRFGGTARLYGEKALQLFADAHICVVGIGGVGSW
                AAEALARTGIGAITLIDMDDVCVTNTNRQIHALRDNVGLAKAEVMAERIRQINSECCV
                TVVDDFVTPDNVAQHMSVGYSYVIDAIDSVRPKAALIAYCRRNKIPLVTTGGAGGQID
                PTQIQVTDLAKTIQDPLAAKLRERLKSDFGVVKNSKGKLGVDCVFSTEALVYPQSDGT
                VCAMKATAEGPKRMDCASGFGAATMVTATFGFVAVSHALKKMMAKAARQG"
gene            complement(3369. .4466)
                /gene="mltA"
                /note="Z4130"
CDS             complement(3369. .4466)
                /gene="mltA"
                /function="enzyme; Cell envelop: Murein sacculus,
                peptidoglycan"
                /note="Residues 1 to 365 of 365 are 100.00 pct identical
                to residues 1 to 365 of 365 from Escherichia coli K-12
                Strain MG1655: B2813"
                /codon_start=1
                /transl_table=11
                /product="membrane-bound lytic murein transglycosylase A"
                /protein_id="AAG57927.1"
                /db_xref="GI:12517295"
                /translation="MKGRWVKYLLMGTVVAMLAACSSKPTDRGQQYKDGKFTQPFSLV
                NQPDAVGAPINAGDFAEQINHIRNSSPRLYGNQSNVYNAVQEWLRAGGDTRNMRQFGI
                DAWQMEGADNYGNVQFTGYYTPVIQARHTRQGEFQYPIYRMPPKRGRLPSRAEIYAGA
                LSDKYILAYSNSLMDNFIMDVQGSGYIDFGDGSPLNFFSYAGKNGHAYRSIGKVLIDR
                GEVKKEDMSMQAIRHWGETHSEAEVRELLEQNPSFVFFKPQSFAPVKGASAVPLVGRA
                SVASDRSIIPPGTTLLAEVPLLDNNGKFNGQYELRLMVALDVGGAIKGQHFDIYQGIG
                PEAGHRAGWYNHYGRVWVLKTAPGAGNVFSG"
gene            4675. .4751
                /gene="metZ"
                /note="Z4131"
tRNA            4675. .4751
                /gene="metZ"
                /product="tRNA-Met"
                /note="anticodon: CAT"
                /function="RNA; tRNA"
gene            4785. .4861
                /gene="metW"
                /note="Z4132"
tRNA            4785. .4861
                /gene="metW"
                /product="tRNA-Met"
                /note="anticodon: CAT"
                /function="RNA; tRNA"
gene            4895. .4971
                /gene="metV"
                /note="Z4133"
tRNA            4895. .4971
                /gene="metV"
                /product="tRNA-Met"
                /note="anticodon: CAT"
                /function="RNA; tRNA"
gene            complement(5045. .6388)
                /gene="Z4134"
CDS             complement(5045. .6388)
                /gene="Z4134"
                /function="putative enzyme; Not classified"
                /note="Residues 1 to 447 of 447 are 99.77 pct identical to
                residues 1 to 447 of 447 from Escherichia coli K-12 Strain
                MG1655: B2817"
                /codon_start=1
                /transl_table=11
                /product="putative amidase"
                /protein_id="AAG57928.1"
                /db_xref="GI:12517296"
                /translation="MTDYASFAKVSGQISRLLVTQLRFLLLGRGMSGSNTAISRRRLL
                QXAGAMWLLSVSQVSLAAVSQVVAVRVWPASSYTRVTVESNRQLKYKQFALXNPERVV
                VDIEDVNLNSVLKGMAAQIRADDPFIKSARVGQFDPQTVRMVFELKQNVKPQLFALAP
                VAGFKERLVMDLYPANAQDMQDPLLALLEDYNKGDLEKQVPPAQSGPQPGKAGRDRPI
                VIMLDPGHGGXDSGAVGKYKTREKDVVLQIARRLRSLIEKEGNMKVYMTRNEDIFIPL
```

```
                                QVRVAKAQKQRADLFVSIHADAFTSRQPSGSSVFALSTKGATSTAAKYLAQTQNASDL
                                IGGVSKSGDRYVDHTMFDMVQSLTIADSLKFGKAVLNKLGKINKLHKNQVEQAGFAVL
                                KAPDIPSILVETAFISNVEEERKLKTATFQQEVAESILAGIKAYFADGATLARRG"
BASE COUNT      1545 a    1761 c    1612 g    1560 t    10 others
ORIGIN

  Query Match           96.7%;  Score 1166.8;  DB 1;  Length 6488;
  Best Local Similarity 97.8%;  Pred. No. 1.6e-291;
  Matches 1180;  Conservative    2;  Mismatches   24;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCCCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
             ||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db     625 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCTCTACAGGATGCGGGC 684

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db     685 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCGA 744

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     745 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 804

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     805 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 864

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     865 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 924

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     925 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 984

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
             |||||||||||||||||||||||||||||| |||||||||||||||||||||| ||||||
Db     985 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTCAAATTG 1044

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
             |||||||||||||||||||||| |||||||||||| |||| ||||||||||||||||||
Db    1045 CCGCTTAATGCGCAGCGACTGCCAGATGTCGATTTGCTGCCGGAACTGATTACTCCCCGT 1104

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
             |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    1105 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTAACTGGCGGTTGCCCGGATCTGGCG 1164

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
             |||||||||||||||||||||||||||||| |||||||| ||||||||||||||||||||
Db    1165 CGAGCGATTACCTTTGCTCATTCAGCCGGGATAGTGGTGATAGTTGATGGTGCTCAGGGG 1224

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
             || |||||||||||||||||||||||||||||| |||||||||||||| || |||
Db    1225 GCGGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATCGATTTCTATGCTTTCTCTGGT 1284

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
             |||||||||||||||||| |||||||||||||||||| |||||||||||||||||||
Db    1285 CACAAACTGTATGGCCCGACGGGTATCGGCGTGCTGTATGGCAAATCAGAACTGCTGGAA 1344

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
             |||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1345 GCGATGTCKCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1404

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
             |||||||||||||||||||||||||||||| |||||||||||||| ||||||||||||
Db    1405 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCAGGAACGCCAAATGTGGCTGGTGTCATA 1464

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
             |||||||||:|||||||||||||||||||||||||||||||||||||||||||| ||||
Db    1465 GGATTAAGCSCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAACTGG 1524

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
             ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db    1525 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCAGGCTTTCGTTCA 1584
```

```
Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db    1585 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCACAGCGAT 1644

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           ||||||||||||||||||||||||||||||||||| |||||||||||| ||||||||||||
Db    1645 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCTTGCGGGCCGGGCAACATTGCGCTCAG 1704

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1705 CCGCTACTGGCAGAATTAGGCGTGACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1764

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1765 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1824

Qy    1201 GATTAA 1206
           ||||||
Db    1825 GATTAA 1830


RESULT 12
AC084742/c
LOCUS       AC084742               251891 bp    DNA     linear   HTG 20-MAY-2002
DEFINITION  Mus musculus chromosome 2 clone RP23-147D3 strain C57BL6/J, WORKING
            DRAFT SEQUENCE, 35 unordered pieces.
ACCESSION   AC084742
VERSION     AC084742.10  GI:20986437
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 251891)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     High Throughput Mouse Sequencing
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 251891)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-NOV-2000) Department of Molecular Genetics, Albert
            Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
            Bronx, NY 10461, USA
REFERENCE   3  (bases 1 to 251891)
  AUTHORS   Grills,G., Han,J., Montgomery,K.T., Chiu,D., Decker,J., Fusina,M.,
            Goltz,J., Haider,A., Hall,L., Ioshikhes,I.P., Lee,E., Long,J.,
            Perera,A., Shim,C., Thomas,E. and Kucherlapati,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-MAY-2002) Harvard Partners Center for Genetics and
            Genomics, Harvard Medical School, 65 Landsdowne St, Cambridge, MA
            02139, USA
COMMENT     On May 20, 2002 this sequence version replaced gi:20219073.
            ----------------Genome Center
            Center:  Harvard Partners Genome Center
            Center Code:  HPGC
            Web site: http://www.hpcgg.org/Sequence/mouse.html
            Contact: hpgc@mendel.mgh.harvard.edu
            ----------------Summary Statistics
            Center project name:  ADK
            Sequencing vector:  pUC18; L08752
            Chemistry:  Dye-terminator Big Dye; 100%
            *Consensus quality:  242636 at least Q20
            *Consensus quality:  237813 at least Q30
            *Consensus quality:  232938 at least Q40
            Estimated insert size:  agarose-FP - N/A
            **Estimated insert size: 251211 - sum-of-contigs
            Quality coverage:  agarose-FP - N/A
            Quality coverage:  7.1 x in Q20 bases; sum-of-contigs estimation
            ----------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 35 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    87943: contig of 87943 bp in length
            *   87944    87963: gap of unknown length
            *   87964   149804: contig of 61841 bp in length
            *  149805   149824: gap of unknown length
            *  149825   176001: contig of 26177 bp in length
            *  176002   176021: gap of unknown length
            *  176022   195216: contig of 19195 bp in length
            *  195217   195236: gap of unknown length
            *  195237   198863: contig of 3627 bp in length
            *  198864   198883: gap of unknown length
            *  198884   199764: contig of 881 bp in length
            *  199765   199784: gap of unknown length
            *  199785   201563: contig of 1779 bp in length
            *  201564   201583: gap of unknown length
            *  201584   204109: contig of 2526 bp in length
            *  204110   204129: gap of unknown length
            *  204130   206919: contig of 2790 bp in length
            *  206920   206939: gap of unknown length
            *  206940   209157: contig of 2218 bp in length
            *  209158   209177: gap of unknown length
            *  209178   211975: contig of 2798 bp in length
            *  211976   211995: gap of unknown length
            *  211996   213476: contig of 1481 bp in length
            *  213477   213496: gap of unknown length
            *  213497   215277: contig of 1781 bp in length
            *  215278   215297: gap of unknown length
            *  215298   217289: contig of 1992 bp in length
            *  217290   217309: gap of unknown length
            *  217310   219571: contig of 2262 bp in length
            *  219572   219591: gap of unknown length
            *  219592   221254: contig of 1663 bp in length
            *  221255   221274: gap of unknown length
            *  221275   223289: contig of 2015 bp in length
            *  223290   223309: gap of unknown length
            *  223310   225551: contig of 2242 bp in length
            *  225552   225571: gap of unknown length
            *  225572   226644: contig of 1073 bp in length
            *  226645   226664: gap of unknown length
            *  226665   228257: contig of 1593 bp in length
            *  228258   228277: gap of unknown length
            *  228278   230116: contig of 1839 bp in length
            *  230117   230136: gap of unknown length
            *  230137   231840: contig of 1704 bp in length
            *  231841   231860: gap of unknown length
            *  231861   233408: contig of 1548 bp in length
            *  233409   233428: gap of unknown length
            *  233429   235021: contig of 1593 bp in length
            *  235022   235041: gap of unknown length
            *  235042   237136: contig of 2095 bp in length
            *  237137   237156: gap of unknown length
            *  237157   238662: contig of 1506 bp in length
            *  238663   238682: gap of unknown length
            *  238683   238772: contig of 90 bp in length
            *  238773   238792: gap of unknown length
            *  238793   239493: contig of 701 bp in length
            *  239494   239513: gap of unknown length
            *  239514   241020: contig of 1507 bp in length
            *  241021   241040: gap of unknown length
            *  241041   243380: contig of 2340 bp in length
            *  243381   243400: gap of unknown length
            *  243401   245028: contig of 1628 bp in length
            *  245029   245048: gap of unknown length
            *  245049   246458: contig of 1410 bp in length
            *  246459   246478: gap of unknown length
```

```
          *   246479   247886: contig of 1408 bp in length
          *   247887   247906: gap of unknown length
          *   247907   249981: contig of 2075 bp in length
          *   249982   250001: gap of unknown length
          *   250002   251891: contig of 1890 bp in length.
FEATURES             Location/Qualifiers
     source          1. .251891
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /chromosome="2"
                     /clone="RP23-147D3"
                     /sex="male"
     misc_feature    1. .87943
                     /note="assembly_name:Contig326
                     clone_end:SP6"
                     vector_side:right"
     misc_feature    87964. .149804
                     /note="assembly_name:Contig325"
     misc_feature    149825. .176001
                     /note="assembly_name:Contig324
                     clone_end:T7
                     vector_side:left"
     misc_feature    176022. .195216
                     /note="assembly_name:Contig323"
     misc_feature    195237. .198863
                     /note="assembly_name:Contig322"
     misc_feature    198884. .199764
                     /note="assembly_name:Contig321"
     misc_feature    199785. .201563
                     /note="assembly_name:Contig320"
     misc_feature    201584. .204109
                     /note="assembly_name:Contig319"
     misc_feature    204130. .206919
                     /note="assembly_name:Contig318"
     misc_feature    206940. .209157
                     /note="assembly_name:Contig317"
     misc_feature    209178. .211975
                     /note="assembly_name:Contig316"
     misc_feature    211996. .213476
                     /note="assembly_name:Contig315"
     misc_feature    213497. .215277
                     /note="assembly_name:Contig314"
     misc_feature    215298. .217289
                     /note="assembly_name:Contig313"
     misc_feature    217310. .219571
                     /note="assembly_name:Contig312"
     misc_feature    219592. .221254
                     /note="assembly_name:Contig311"
     misc_feature    221275. .223289
                     /note="assembly_name:Contig310"
     misc_feature    223310. .225551
                     /note="assembly_name:Contig309"
     misc_feature    225572. .226644
                     /note="assembly_name:Contig308"
     misc_feature    226665. .228257
                     /note="assembly_name:Contig307"
     misc_feature    228278. .230116
                     /note="assembly_name:Contig306"
     misc_feature    230137. .231840
                     /note="assembly_name:Contig305"
     misc_feature    231861. .233408
                     /note="assembly_name:Contig304"
     misc_feature    233429. .235021
                     /note="assembly_name:Contig303"
     misc_feature    235042. .237136
                     /note="assembly_name:Contig302"
     misc_feature    237157. .238662
                     /note="assembly_name:Contig301"
     misc_feature    238683. .238772
                     /note="assembly_name:Contig300"
     misc_feature    238793. .239493
                     /note="assembly_name:Contig299"
     misc_feature    239514. .241020
                     /note="assembly_name:Contig298"
     misc_feature    241041. .243380
                     /note="assembly_name:Contig297"
     misc_feature    243401. .245028

  Query Match             77.9%;  Score 939;  DB 2;  Length 251891;
  Best Local Similarity   100.0%;  Pred. No. 2.3e-232;
  Matches  939;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCCCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247435 ATGAACGTTTTTAATCCCGCGCAGTTTCCCGCCCAGTTTCCCGCACTACAGGATGCGGGC 247376

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247375 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 247316

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247315 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 247256

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247255 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 247196

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247195 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 247136

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247135 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 247076

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247075 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 247016

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 247015 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 246956

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246955 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 246896

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246895 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 246836

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246835 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 246776

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246775 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 246716

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246715 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 246656

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246655 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 246596

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 246595 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 246536

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCG 939
          |||||||||||||||||||||||||||||||||||||||
Db 246535 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCG 246497
```

```
RESULT 13
AE008836
LOCUS       AE008836               20653 bp    DNA     linear   BCT 31-JUL-2002
DEFINITION  Salmonella typhimurium LT2, section 140 of 220 of the complete
            genome.
ACCESSION   AE008836 AE006468
VERSION     AE008836.1  GI:16421517
KEYWORDS    .
SOURCE      Salmonella typhimurium LT2.
  ORGANISM  Salmonella typhimurium LT2
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 20653)
  AUTHORS   McClelland,M., Sanderson,K.E., Spieth,J., Clifton,S.W.,
            Latreille,P., Courtney,L., Porwollik,S., Ali,J., Dante,M., Du,F.,
            Hou,S., Layman,D., Leonard,S., Nguyen,C., Scott,K., Holmes,A.,
            Grewal,N., Mulvaney,E., Ryan,E., Sun,H., Florea,L., Miller,W.,
            Stoneking,T., Nhan,M., Waterston,R. and Wilson,R.K.
  TITLE     Complete genome sequence of Salmonella enterica serovar Typhimurium
            LT2
  JOURNAL   Nature 413 (6858), 852-856 (2001)
  MEDLINE   21534948
   PUBMED   11677609
REFERENCE   2  (bases 1 to 20653)
  AUTHORS   The Salmonella typhimurium Genome Sequencing Project.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAR-2001) Genome Sequencing Center, Department of
            Genetics, Washington University School of Medicine, 4444 Forest
            Park Boulevard, St. Louis, MO 63108, USA
COMMENT     COMMENT     Supported by NIH grant 5U 01 AI43283

            Coding sequences below are predicted from manually evaluated
            computer analysis, using similarity information and the programs;
            GLIMMER; http://www.tigr.org/softlab/glimmer/glimmer.html and
            GeneMark; http://opal.biology.gatech.edu/GeneMark/

            EC numbers were kindly provided by Junko Yabuzaki and the Kyoto
            Encyclopedia of Genes and Genomes; http://www.genome.ad.jp/kegg/,
            and Pedro Romero and Peter Karp at EcoCyc;
            http://ecocyc.PangeaSystems.com/ecocyc/

            The analyses of ribosome binding sites and promoter binding sites
            were kindly provided by Heladia Salgado, Julio Collado-Vides and
            ReguonDB;
            http://kinich.cifn.unam.mx:8850/db/regulondb_intro.frameset

            This sequence was finished as follows unless otherwise noted: all
            regions were double stranded, sequenced with an alternate
            chemistries or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one m13 subclone.
FEATURES             Location/Qualifiers
     source          1. .20653
                     /organism="Salmonella typhimurium LT2"
                     /strain="LT2; SGSC 1412; ATCC 700720"
                     /db_xref="ATCC:700720"
                     /db_xref="taxon:99287"
                     /note="LT2"
     gene            complement(139. .693)
                     /gene="syd"
                     /note="STM2967"
     CDS             complement(139. .684)
                     /gene="syd"
                     /note="similar to E. coli interacts with secY
                     (AAC75835.1); Blastp hit to AAC75835.1 (181 aa), 83%
                     identity in aa 1 - 181"
                     /codon_start=1
                     /transl_table=11
                     /product="interacts with secY"
                     /protein_id="AAL21846.1"
                     /db_xref="GI:16421518"
                     /translation="MDELTAQALKAFTTRYCDAWQEKHGSWPLSEELYGVPSPCIISS
                     TRDAVYWQPQPFEGEENVNAVERAFDIMVQPALHAFYTTQFAGDMPAQFADEKLTLLQ
                     TWSQDDFRRVQENLIGHLVTQKRLKLPPTLFIATQENELEVISVCNLSGEVIKETLGT
                     RNRTVLAATLAEFLTQLNPLL"
RBS                  complement(688. .693)
                     /gene="syd"
                     /note="putative RBS for syd; RegulonDB:STMS1H003019"
gene                 733. .1601
                     /gene="yqcD"
                     /note="STM2968"
RBS                  733. .738
                     /gene="yqcD"
                     /note="putative RBS for yqcD; RegulonDB:STMS1H003020"
CDS                  753. .1601
                     /gene="yqcD"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC75836.1); Blastp hit to AAC75836.1 (282 aa), 87%
                     identity in aa 1 - 282"
                     /codon_start=1
                     /transl_table=11
                     /product="putative GTP cyclohydrolase I"
                     /protein_id="AAL21847.1"
                     /db_xref="GI:16421519"
                     /translation="MSSYENHQALDGLTLGKSTDYRDNYDVSLLQGVPRSLNRDPLGL
                     TADNLPFHGADIWTLYELSWLNSQGLPQVAVGHVELDYTSVNLIESKSFKLYLNSFNQ
                     TRFDTWETVRQTLERDLRACAQGNVSVRLHRLDELEGQPVAHFHGTCIDDQDISIDNY
                     QFTTDYLQHAVSGEKQVEETLVSHLLKSNCLITHQPDWGSIQIQYRGRKIDREKLLRY
                     LVSFRHHNEFHEQCVERIFNDILRFCQPETLSVYARYTRRGGLDINPWRSNTDFVPAT
                     GRLARQ"
gene                 1703. .3078
                     /gene="ygdH"
                     /note="STM2969"
RBS                  1703. .1708
                     /gene="ygdH"
                     /note="putative RBS for ygdH; RegulonDB:STMS1H003021"
CDS                  1714. .3078
                     /gene="ygdH"
                     /note="similar to E. coli orf, hypothetical protein
                     (AAC75837.1); Blastp hit to AAC75837.1 (454 aa), 94%
                     identity in aa 1 - 453"
                     /codon_start=1
                     /transl_table=11
                     /product="putative nucleotide binding"
                     /protein_id="AAL21848.1"
                     /db_xref="GI:16421520"
                     /translation="MITHISPLGSMDMLSQLEVDMLKRTASSDLYQLFRNCSLAVLNS
                     GSLTDNSKELLSRFENFDINVLRRERGVKLELINPPEDAFVDGRIIRALQANLFAVLR
                     DILFVNGQIHNAGRFQHLDLESSTHITNLVFSILRNARALHVGEAPNMVVCWGGHSIN
                     ENEYLYARRVGTQLGLRELNICTGCGPGAMEAPMKGAAVGHAQQRYKDSRFIGMTEPS
                     IIAAEPPNPLVNELIIMPDIEKRLEAFVRIAHGIIIFPGGVGTAEELLYLLGILMNPA
                     NKNQVLPLILTGPKESADYFRVLDEFITHTLGEAARRHYRIIIDDAAEVARLMKKAMP
                     QVKENRRDTGDAYSFNWSMRIAPDLQVPFEPSHENMANLKLYPDQPVEILAADLRRAF
                     SGIVAGNVKEVGIRAIEANGPYKIHGDREMMRRMDDLLQGFVAQHRMKLPGSAYIPCY
                     EICA"
gene                 3631. .4932
                     /gene="sdaC"
                     /note="STM2970"
RBS                  3631. .3636
                     /gene="sdaC"
                     /note="putative RBS for sdaC; RegulonDB:STMS1H003022"
CDS                  3643. .4932
                     /gene="sdaC"
                     /note="similar to E. coli probable serine transporter
                     (AAC75838.1); Blastp hit to AAC75838.1 (429 aa), 95%
                     identity in aa 1 - 429"
                     /codon_start=1
                     /transl_table=11
                     /product="putative HAAAP family serine transport protein"
                     /protein_id="AAL21849.1"
                     /db_xref="GI:16421521"
                     /translation="METTQTSTIASIDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPIN
                     AGVGGMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLIT
```

```
                LLYFFAIYPILLVYSVAITNTVESFLTHQLAINPPPRAILSLILIVGMMTIVRFGEQM
                IVKAMSILVFPFVAALMLLALYLIPQWNGAALETLSFDSAASTGNGLWMTLWLAIPVM
                VFSFNHSPIISSFAVAKREEYGEGAEKKCSKILAFAHIMMVLTVMFFVFSCVLSLTPA
                DLAAAKEQNISILSYLANHFNAPIIAWMAPIIAMIAITKSFLGHYLGAREGFNGMVIK
                SLRGKGKSIEINKLNKITALFMLVTTWIVATLNPSILGMIETLGGPIIAMILFLMPMY
                AIQKVPAMRKYSGHISNVFVVIMGLIAISAIFYSLFS"
     gene       4980. .6358
                /gene="sdaB"
                /note="STM2971"
     RBS        4980. .4985
                /gene="sdaB"
                /note="putative RBS for sdaB; RegulonDB:STMS1H003023"
     CDS        4991. .6358
                /gene="sdaB"
                /EC_number="4.2.1.13"
                /note="similar to E. coli L-serine dehydratase
                (deaminase), L-SD2 (AAC75839.1); Blastp hit to AAC75839.1
                (455 aa), 94% identity in aa 1 - 455; L-threonine
                deaminase 2"
                /codon_start=1
                /transl_table=11
                /product="L-serine dehydratase"
                /protein_id="AAL21850.1"
                /db_xref="GI:16421522"
                /translation="MISVFDIFKIGIGPSSSHTVGPMKAGKQFTDDLIARHILTDVTR
                VVVDVYGSLSLTGKGHHTDIAIIMGLAGNLPDTVDIDAIPSFIQDVNTHGRLLLANGQ
                HEVEFPVDKCMNFHADNLSLHENGMRITALAGDKVLYSQTYYSIGGGFIVDEEHFGLT
                NSEPVNVPYPYKTAADLQRHCQETGLSLSGLMMQNELALHSKEALEQHFARVWEVMRS
                GIERGITTEGVLPGKLRVPRRAAALRRMLVSQDKTTTDPMAVVDWINMFALAVNEENA
                AGGRVVTAPTNGACGIVPAVLAYYDKFIREVNANSLARYMLVASAIGSLYKMNASISG
                AEVGCQGEVGVACSMAAAGLAELLGGSPAQVCIAAEIAMEHNLGLTCDPVAGQVQVPC
                IERNAIASVKAVNAARMAVRRTSEPRVCLDKVIETMYETGKDMNAKYRETSRGGLAMK
                IVTCD"
     gene       6449. .7284
                /gene="exo"
                /note="STM2972"
     RBS        6449. .6454
                /gene="exo"
                /note="putative RBS for exo; RegulonDB:STMS1H003024"
     CDS        6469. .7284
                /gene="exo"
                /EC_number="3.1.11.-"
                /note="similar to E. coli 5'-3' exonuclease (AAC75840.1);
                Blastp hit to AAC75840.1 (281 aa), 87% identity in aa 18 -
                281"
                /codon_start=1
                /transl_table=11
                /product="exonuclease IX, 5'-3' exonuclease"
                /protein_id="AAL21851.1"
                /db_xref="GI:16421523"
                /translation="MLASPRRECYPTRRFVRENTVAAHLLIVDALNLIRRIHAVQGSP
                CVETCQHALDQLIIHSQPTHAVAVFDDDARSSGWRHQRLPDYKAGRPPMPDDLHNEMP
                ALRAAFEQRGVRCWASDGNEADDLAATLALKVTEAGHQATIVSTDKGYCQLLSPGLRI
                RDYFQKRWLDAPFIEKEFGVLPRQLPDYWGLAGISSSKVPGVAGIGPKSATQLLIQFQ
                NLEGIYAHLDEVPEKWRKKLETHKEMAFLCRDIARLQTDLHIDGNLQQLRLAR"
     gene       complement(7376. .8537)
                /gene="fucO"
                /note="STM2973"

 Query Match             69.1%;  Score 833.2;  DB 1;  Length 20653;
 Best Local Similarity   80.7%;  Pred. No. 5.3e-205;
 Matches  973;  Conservative    0;  Mismatches  233;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
           ||||||| |||||||||| |||||||||||| |||||||||| |||   ||||||||
Db   18828 ATGAACGCTTTTAATCCCACGCAGTTTCGCGCGCAGTTTCCCGCGCTAGCCGATGCGGGT 18887

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           || |||||||| |||||||| || ||  | || ||  | || ||  |||| ||||| ||
Db   18888 GTTTATCTCGATAGCGCCGCCACGGCATTAAAGCCACAGGCAGTCATTGACGCCACGCAC 18947

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
           ||||||||  || |||||||||| ||||| ||||| |||||||||||  | || || |||

Db   18948 CAGTTTTATTGTTTGAGCGCCGGTAACGTTCATCGTAGCCAGTTTGCGCAGGCGCAGCGC 19007

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           ||||| |||| |||||||| || ||| |||| |||| | | || || || |||
Db   19008 CTGACGGCGCAATATGAAGCGGCCAGAGCAAAAGCAGCGCGGCTGTTAAACGCGCCCGAT 19067

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
           || |||| |||||||||||| |||||||||| ||| | |||||||||||||| || ||
Db   19068 GAAAAAAGTATCGTCTGGACACGCGGCACCACCGAAGCGATCAACATGGTGGCGCAGTGT 19127

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
           || || ||||| ||||||| || |||||||| |||| || ||||| || || || |||
Db   19128 TACGCCCGTCCTCGTCTGCGCCCCGGCGATGAAATTATCGTTAGCGTCGCCGAGCATCAC 19187

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           |||||||| || |||||||||||||| || |||||||| || ||  | ||  | ||| ||
Db   19188 GCCAACCTTGTGCCCTGGCTGATGGTGGCGCAACAAACCGGCGCGCAGGTCATAAAACTG 19247

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
           |||||||||  | ||| || || ||||| ||     |||| |||||||| ||  | ||
Db   19248 CCGCTTAATGACCGGCGTCTTCCTGATGTTGAGCGTCTGCCGGAACTGATCACGTCGCGC 19307

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
           ||| |||||||||||| |||| || |||||||||| || |||| ||||||||||| ||
Db   19308 AGCCGGATTCTGGCGCTGGGGCAAATGTCGAACGTAACGGGCGGCTGCCCGGATCTCGCA 19367

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
           | || || | |  |||||  |||| || ||||| |||||||| ||||| || || ||
Db   19368 AGCGCTATCAGCGCCGCTCACGCAGCGGGAATGGTCGTGATGGTAGATGGCGCGCAAGGC 19427

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
           || || || ||||| |||||||||||||| || ||||| ||||| |||||||||| | |
Db   19428 GCGGTACACTTCCCGGCGGATGTTCAGCAGCTTGATATCGATTTTTATGCTTTTTCCGCT 19487

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           |||||||||||||||||| |||||||||||||||| ||||| | || || || |||
Db   19488 CACAAACTGTATGGCCCGACCGGTATCGGCGTGCTGTACGGTAAGCCGGAGCTTCTTGAG 19547

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
           |||||||||||||||| |||||||||| ||| | |  || || || ||||| ||||||
Db   19548 GCGATGTCGCCCTGGCTCGGCGGCGGCAAGATGATCCGTGACGTTAGCTTTGAAGGCTTC 19607

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           || ||||||   || || |||||||||||||| || ||||| || ||||| || |||||
Db   19608 ACCACTCAAAGCGCTCCCTGGAAACTGGAAGCGGGGACGCCGAACGTCGCCGGGGTCATC 19667

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           ||  | ||||| |||||||||||||| | |||  |||||  | ||||||||| ||||
Db   19668 GGCCTGAGCGCTGCGCTGGAATGGCTGTCCGATATCGATATTGAACAGGCCGAAAACTGG 19727

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
           |||||  |  | || |||||||||||||| || |||||||||| || ||||||||||
Db   19728 AGCCGCGGGCTGGCGACGCTGGCGGAAGACGCACTGGCGAAACGCCCGGGCTTTCGTTCG 19787

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
           |||||||||||||| |||||||||||||||||||||||||  ||||| || ||  |||||
Db   19788 TTCCGCTGCCAGGACTCCAGCCTGCTGGCCTTTGATTTTGTCGGCGTGCACCACGGCGAT 19847

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
           |||||||||||||||||| |||||||||| ||||||||||| ||||||||| |||||||
Db   19848 ATGGTGACGCTGCTGGCGGAATACGGTATTGCGCTCCGGGCCGGGCAACATTGCGCCCAG 19907

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
           ||  | ||||| ||| | |||| || || || ||||||||||||||||| ||||||
Db   19908 CCATTGCTGGCGGAACTTGGCGTCACAGGGACTCTGCGCGCCTCTTTTGCGCCGTATAAT 19967

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCGCTGGAATTATTGGTG 1200
           ||  |  |||||||||||||| |||| || ||||||||||||||||||| |  |||||
Db   19968 ACCCAACATGATGTGGATGCGTTGGTTAACGCCGTTGACCGCGCGCGCTGGAACTGCTGGTG 20027

Qy    1201 GATTAA 1206
           ||||||
Db   20028 GATTAA 20033
```

```
RESULT 14
AL627277
LOCUS       AL627277              230050 bp    DNA     linear   BCT 06-JUN-2002
DEFINITION  Salmonella enterica serovar Typhi (Salmonella typhi) strain CT18,
            complete chromosome; segment 13/20.
ACCESSION   AL627277 AL513382
VERSION     AL627277.1  GI:16504049
KEYWORDS    .
SOURCE      Salmonella enterica subsp. enterica serovar Typhi.
  ORGANISM  Salmonella enterica subsp. enterica serovar Typhi
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 230050)
  AUTHORS   Parkhill,J., Dougan,G., James,K.D., Thomson,N.R., Pickard,D.,
            Wain,J., Churcher,C., Mungall,K.L., Bentley,S.D., Holden,M.T.G.,
            Sebaihia,M., Baker,S., Basham,D., Brooks,K., Chillingworth,T.,
            Connerton,P., Cronin,A., Davis,P., Davies,R.M., Dowd,L., White,N.,
            Farrar,J., Feltwell,T., Hamlin,N., Haque,A., Hien,T.T., Holroyd,S.,
            Jagels,K., Krogh,A., Larsen,T.S., Leather,S., Moule,S., O'Gaora,P.,
            Parry,C., Quail,M., Rutherford,K., Simmonds,M., Skelton,J.,
            Stevens,K., Whitehead,S. and Barrell,B.G.
  TITLE     Complete genome sequence of a multiple drug resistant Salmonella
            enterica serovar Typhi CT18
  JOURNAL   Nature 413 (6858), 848-852 (2001)
  MEDLINE   21534947
   PUBMED   11677608
REFERENCE   2  (bases 1 to 230050)
  AUTHORS   Parkhill,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-OCT-2001) Submitted on behalf of the Salmonalla
            sequencing team, Sanger Centre, Wellcome Trust Genome Campus,
            Hinxton, Cambridge CB10 1SA, UK
COMMENT     E-mail: parkhill@sanger.ac.uk
            Notes:
            Details of S. typhi sequencing at the Sanger Centre are available
            on the World Wide Web.
            (URL, http://www.sanger.ac.uk/Projects/S_typhi/).
FEATURES             Location/Qualifiers
     source          1. .230050
                     /organism="Salmonella enterica subsp. enterica serovar
                     Typhi"
                     /strain="CT18"
                     /db_xref="taxon:90370"
     gene            309. .1598
                     /gene="STY3109"
                     /note="sdaC"
     CDS             309. .1598
                     /gene="STY3109"
                     /note="Fasta hit to Q46940 (409 aa), 40% identity in 416
                     aa overlap
                     Fasta hit to TDCC_ECOLI (443 aa), 50% identity in 442 aa
                     overlap
                     Orthologue of E. coli sdaC (SDAC_ECOLI); Fasta hit to
                     SDAC_ECOLI (429 aa), 95% identity in 429 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="putative serine transporter"
                     /protein_id="CAD02796.1"
                     /db_xref="GI:16504050"
                     /db_xref="SPTREMBL:Q8Z435"
                     /translation="MKTTQTSTIASIDSRSAWRKTDTMWMLGLYGTAIGAGVLFLPIN
                     AGVGGMIPLIIMAILAFPMTFFAHRGLTRFVLSGKNPGEDITEVVEEHFGIGAGKLIT
                     LLYFFAIYPILLVYSVAITNTVESFLTHQLAINPPPRAILSLILIVGMMTIVRFGEQM
                     IVKAMSILVFPFVAALMLLALYLIPQWNGAALETLSFDSAASTGNGLWMTLWLAIPVM
                     VFSFNHSPIISSFAVAKREEYGEGAEKKCSKILAFAHIMMVLTVMFFVFSCVLSLTPA
                     DLAAAKEQNISILSYLANHFNAPIIAWMAPIIAMIAITKSFLGHYLGAREGFNGMVIK
                     SLRGKGKSIEINKLNKITALFMLVTTWIVATLNPSILGMIETLGGPIIAMILFLMPMY
                     AIQKVPAMRKYSGHISNVFVVIMGLIAISAIFYSLFS"
     gene            1657. .3024
                     /gene="STY3110"
                     /note="sdaB"
     CDS             1657. .3024
                     /gene="STY3110"
                     /note="Fasta hit to TDCG_ECOLI (454 aa), 73% identity in
                     454 aa overlap
                     Fasta hit to SDHL_ECOLI (454 aa), 77% identity in 455 aa
                     overlap
                     Orthologue of E. coli sdaB (SDHM_ECOLI); Fasta hit to
                     SDHM_ECOLI (455 aa), 94% identity in 455 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="L-serine dehydratase 2 (L-serine deaminase 2)"
                     /protein_id="CAD02797.1"
                     /db_xref="GI:16504051"
                     /db_xref="SPTREMBL:Q8Z434"
                     /translation="MISVFDIFKIGIGPSSSHTVGPMKAGKQFTDDLIARHILTDVTR
                     VVVDVYGSLSLTGKGHHTDIAIIMGLAGNLPDTVDIDAIPSFIQDVNTHGRLLLANGQ
                     HEVEFPVDKCMNFHADNLSLHENGMRITALAGNKVLYSQTYYSIGGGFIVDEEHFGLT
                     NSEPVNVPYPYKTAADLQRHCQETGLSLSGLMMQNELALHSKEALEQHFAKVWEVMRS
                     GIERGITTEGVLPGKLRVPRRAAALRRMLVSQDKTTTDPMAVVDWINMFALAVNEENA
                     AGGRVVTAPTNGACGIVPAVLAYYDKFIREVNANSLARYMLVASAIGSLYKMNASISG
                     AEVGCQGEVGVACSMAAAGLAELLGGSPAQVCIAAEIAMEHNLGLTCDPVAGQVQVPC
                     IERNAIASVKAVNAARMALRRTSEPRVCLDKVIETMYETGKDMNAKYRETSRGGLAMK
                     IVTCD"
     gene            3135. .3950
                     /gene="STY3111"
                     /note="exo"
     CDS             3135. .3950
                     /gene="STY3111"
                     /note="Orthologue of E. coli exo (EX9_ECOLI); Fasta hit to
                     EX9_ECOLI (281 aa), 88% identity in 264 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="5'-3' exonuclease"
                     /protein_id="CAD02798.1"
                     /db_xref="GI:16504052"
                     /db_xref="SPTREMBL:Q8Z433"
                     /translation="MLASPRRECYPTRRFVRENTVAAHLLIVDALNLIRRIHAVQGSP
                     CVETCQHALDQLIIHSQPTHAVAVFDDDARSSGWRHQRLPDYKAGRPPMPDDLHNEMP
                     ALRAAFEQRGVRCWASDGNEADDLAATLALKVTEAGHQATIVSTDKGYCQLLSPGLRI
                     RDYFQKRWLDAPFIEKEFGVLPRQLPDYWGLAGISSSKVPGVAGIGPKSATQLLVQFQ
                     NLEGIYAHLDEVPEKWRKKLETHKEMAFLCRDIARLQTDLHIDGNLQQLRLVR"
     misc_feature    3201. .3947
                     /gene="STY3111"
                     /note="Pfam match to entry PF01367 5_3_exonuclease, 5'-3'
                     exonuclease, score 120.40, E-value 3.3e-32"
     gene            complement(4042. .5190)
                     /gene="STY3112"
                     /note="fucO"
     CDS             complement(4042. .5190)
                     /gene="STY3112"
                     /note="Fasta hit to EUTG_ECOLI (395 aa), 40% identity in
                     372 aa overlap
                     Fasta hit to ADH2_ECOLI (383 aa), 44% identity in 371 aa
                     overlap
                     Orthologue of E. coli fucO (FUCO_ECOLI); Fasta hit to
                     FUCO_ECOLI (383 aa), 91% identity in 382 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="1,2-propanediol oxidoreductase (lactaldehyde
                     reductase)"
                     /protein_id="CAD02799.1"
                     /db_xref="GI:16504053"
                     /db_xref="SPTREMBL:Q8Z432"
                     /translation="MANRMILNETAWFGRGAVDALTDEVTRRGYHKALIVTDKTLVQC
                     GVVDKVTSRMDAAGLAWEIYAGVIPNPTISVVQEGLKVFTQSGADYLIAIGGGSPQDT
                     CKAVGIISNNPEFADVRSLEGLSPTRKPSVPIMAIPTTAGTAAEVTINYVITDEEKRR
                     KFVCVDPHDIPQAAFIDADMMDGMPPALKAATGVDALTHAIEGYITRAAWALTDALHI
                     KAIEIIAGALRGAVAGEKEAGEAMALGQYVAGMGFSNVGLGLVHGMAHPLGAFYNTPH
                     GVANAILLPHVMRFNAGSTNEKFRDIARAMGVKVEGLSLEEARNAAVEAVFTLNRDVG
                     IPLHLRDVGVRKEDIPALAQAAFDDVCTGGNPREASLADIVELYHTAW"
     misc_feature    complement(4069. .5169)
                     /gene="STY3112"
                     /note="Pfam match to entry PF00465 Fe-ADH, Iron-containing
```

```
                alcohol dehydrogenases, score 655.70, E-value 2.4e-193"
     misc_feature    complement(4354. .4416)
                     /gene="STY3112"
                     /note="PS00060 Iron-containing alcohol dehydrogenases
                     signature 2"
     misc_feature    complement(4585. .4671)
                     /gene="STY3112"
                     /note="PS00913 Iron-containing alcohol dehydrogenases
                     signature 1"
     gene            complement(5207. .5854)
                     /gene="STY3113"
                     /note="fucA"
     CDS             complement(5207. .5854)
                     /gene="STY3113"
                     /note="Fasta hit to YGBL_ECOLI (212 aa), 34% identity in
                     179 aa overlap
                     Orthologue of E. coli fucA (FUCA_ECOLI); Fasta hit to
                     FUCA_ECOLI (215 aa), 92% identity in 215 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="fuculose-1-phosphate aldolase"
                     /protein_id="CAD02800.1"
                     /db_xref="GI:16504054"
                     /db_xref="SPTREMBL:Q8Z431"
                     /translation="MERNRLARQIIDTCLEMTRLGLNQGTAGNVSVRYQGGMLITPTG
                     IPYEKLTESHIVFIDADGQHEQGKLPSSEWRFHMAAYQTRPDANAVVHNHAVHCTAVS
                     ILNRPIPAIHYMIAAAGGNSIPCAPYATFGTRELSDHVAVALKNRKATLLQHHGLIAC
                     EENLDKALWLAHEVEVLAQLYLSTLAIVDPVPVLDDEAIAIVLEKFKTYGLRIEE"
     misc_feature    complement(5219. .5845)
                     /gene="STY3113"
                     /note="Pfam match to entry PF00596 Aldolase_II, Class II
                     Aldolase and Adducin N-terminal domain., score 337.10,
                     E-value 2e-97"
     gene            6410. .7726
                     /gene="STY3115"
                     /note="fucP"
     CDS             6410. .7726
                     /gene="STY3115"
                     /note="Orthologue of E. coli fucP (FUCP_ECOLI); Fasta hit
                     to FUCP_ECOLI (438 aa), 92% identity in 438 aa overlap"
                     /codon_start=1
                     /transl_table=11
                     /product="L-fucose permease"
                     /protein_id="CAD02801.1"
                     /db_xref="GI:16504055"
                     /db_xref="SPTREMBL:Q8Z430"
                     /translation="MGNTTIQTQSFRAVDAEQSKSKRYIIPFALLCSLFFLWAVANNL
                     NDILLPQFQQAFTLTNFQAGLIQSAFYFGYFVIPIPAGILMKKLSYKAGIITGLFLYA
                     FGAALFWPAAEIMNYTLFLIGLFIIAAGLGCLETAANPFVTALGPESGGHFRLNLAQT
                     FNSFGAIIAVVFGQSLILSNVPHQSQEALDKMTPDQLSAYKHSLVLSVQTPYMIIVAI
                     VLVVALLIMLTKFPALQSDDHSDAKQSSFLSSLSRLIRIRHWRWAVLAQFCYVGAQTA
                     CWSYLIRYAIEEIPGMTPGFAANYLTGTMVCFFIGRFTGTWLISRFAPHKVLAAYALL
                     AMLLCLISAFSGGHIGLLALTLCSAFMSIQYPTIFSLGIKNLGQDTKYGSSFIVMTII
                     GGGIVTPVMGFVSDAAGKIPTAELVPALCFAVIFIFARFRSQAATN"
     gene            7758. .9488
                     /gene="STY3116"
                     /note="fucI"
     CDS             7758. .9488

  Query Match          68.3%;  Score 823.6;  DB 1;  Length 230050;
  Best Local Similarity  80.2%;  Pred. No. 1.9e-202;
  Matches  967;  Conservative   0;  Mismatches  239;  Indels    0;  Gaps    0;

Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||| |||||||||| |||||||||||| |||||||||| |||   ||||||||
Db  15494 ATGAACGCTTTTAATCCCACGCAGTTTCGCGCGCAGTTTCCCGCGCTAGCCGATGCGGGT 15553

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          || |||||||| |||||||| || ||  | || ||  | || ||  |||| ||||| ||
Db  15554 GTTTATCTCGATAGCGCCGCCACGGCATTAAAGCCACAGGCAGTCATTGACGCCACGCAC 15613

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||  || ||||||||| ||||| ||||| |||||||||||  | || || |||

Db  15614 CAGTTTTATTGTTTGAGCGCCGGTAACGTTCATCGTAGCCAGTTTGCGCAGGCGCAGCGC 15673

Qy    181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||| |||| |||||||| ||  |||  |||| |||||   |  | || || || |||
Db  15674 CTGACGGCGCAATATGAAGCGGCCAGAGCAAAAGCGGCGCGGCTGTTAAACGCGCCCGAT 15733

Qy    241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          || |||| |||||||||||| |||||||||||| ||| | ||||||||||||||| || ||
Db  15734 GAAAAAAGTATCGTCTGGACACGCGGCACCACCGAAGCGATCAACATGGTGGCGCAGTGT 15793

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          || || ||||| ||||||| || |||||||| ||||| || ||||| || || || |||
Db  15794 TACGCCCGTCCTCGTCTGCGCCCCGGCGATGAAATTATCGTTAGCGTCGCCGAGCATCAC 15853

Qy    361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||| || |||||||||||||| || ||||||| || ||  | ||  | ||| ||
Db  15854 GCCAACCTTGTTCCCTGGCTGATGGTGGCGCAACAAACCGGCGCGCAGGTCATAAAACTG 15913

Qy    421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          |||||||||  |  || || || ||||| ||     |||| |||||||| ||  | ||
Db  15914 CCGCTTAATGACCGACGTCTTCCTGATGTTGAGCGTCTGCCGGAACTGATCACGTCGCGC 15973

Qy    481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          || ||||| ||||| |||| || |||||||||| || ||||| ||||| |||||||| |
Db  15974 AGCCGGATACTGGCGCTGGGGCAAATGTCGAACGTAACGGGCGGCTGTCCGGATCTCACA 16033

Qy    541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          | || || | |  ||||| | || || ||||| |||||| | ||||| || || ||
Db  16034 GGCGCTATCAGCGCCGCTCACGCGGCGGGAATGGTCGTGATGATAGATGGCGCGCAAGGC 16093

Qy    601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          || || || ||||| |||||||| |||| || ||||| ||||| |||||||||| | |
Db  16094 GCTGTACACTTCCCGGCGGATGTCCAGCAGCTTGATATCGATTTTTATGCTTTTTCCGCT 16153

Qy    661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          |||||||||||||||||| |||||||||||||||| |||||  | || || || |||
Db  16154 CACAAACTGTATGGCCCGACCGGTATCGGCGTGCTGTACGGTAAGCCGGAGCTTCTTGAG 16213

Qy    721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          |||||||||||||||| |||||||||| ||| | |  || || || ||||| ||||||
Db  16214 GCGATGTCGCCCTGGCTCGGCGGCGGCAAGATGATCCGTGACGTTAGCTTTGAAGGCTTC 16273

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          || ||||||   || || |||||||||||||| || ||||| || ||||| || |||||
Db  16274 ACCACTCAAAGCGCTCCCTGGAAACTGGAAGCGGGGACGCCGAACGTCGCCGGGGTCATC 16333

Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||  | ||||| |||||||||||||| | |||  ||||||  | ||||||||| ||||
Db  16334 GGCCTGAGCGCTGCGCTGGAATGGCTGTCCGATATCGATATTGAACAGGCCGAAAACTGG 16393

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          |||||  |  | || |||||||||||||| || ||||||||||||| ||||||||||
Db  16394 AGCCGGGGGCTGGCGACGCTGGCGGAAGACGCACTGGCGAAACGTCCGGGCTTTCGTTCG 16453

Qy    961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          |||||||||||||| |||||||||||||||||||||||||  |||| || ||  |||||
Db  16454 TTCCGCTGCCAGGACTCCAGCCTGCTGGCCTTTGATTTTGTCGGCGTGCACCACGGCGAT 16513

Qy   1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          |||||||||||||||||| |||||||||| || |||||||||| |||||||| |||
Db  16514 ATGGTGACGCTGCTGGCGGAATACGGTATTGCGCTCCGGGCCGGGCAACATTGCGCCCAG 16573

Qy   1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||  | || || ||| | ||||| || ||  | ||||||||||||||||||| ||||||
Db  16574 CCATTGCTAGCGGAACTTGGCGTCACAGGGACTCTGCGCGCCTCTTTTGCGCCGTATAAT 16633

Qy   1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCGCTGGAATTATTGGTG 1200
          ||  |  |||||||||||||| |||| || ||||||||||||||||||| |  |||||
Db  16634 ACCCAACATGATGTGGATGCGTTGGTTAACGCCGTTGACCGCGCGCGCTGGAACTGCTGGTG 16693

Qy   1201 GATTAA 1206
          ||||||
Db  16694 GATTAA 16699
```

```
RESULT 15
AE013916
LOCUS       AE013916               9280 bp    DNA     linear   BCT 26-JUL-2002
DEFINITION  Yersinia pestis KIM section 316 of 415 of the complete genome.
ACCESSION   AE013916 AE009952
VERSION     AE013916.1  GI:21960078
KEYWORDS    .
SOURCE      Yersinia pestis KIM.
  ORGANISM  Yersinia pestis KIM
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Yersinia.
REFERENCE   1  (bases 1 to 9280)
  AUTHORS   Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
            Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
            Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
            Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
            Blattner,F.R. and Perry,R.D.
  TITLE     Genome Sequence of Yersinia pestis KIM
  JOURNAL   J. Bacteriol. 184 (16), 4601-4611 (2002)
   PUBMED   12142430
REFERENCE   2  (bases 1 to 9280)
  AUTHORS   Deng,W., Burland,V., Plunkett,G. III, Boutin,A., Mayhew,G.F.,
            Liss,P., Perna,N.T., Rose,D.J., Mau,B., Zhou,S., Schwartz,D.C.,
            Fetherston,J.D., Lindler,L.E., Brubaker,R.R., Plana,G.V.,
            Straley,S.C., McDonough,K.A., Nilles,M.L., Matson,J.S.,
            Blattner,F.R. and Perry,R.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-FEB-2002) Genetics, University of Wisconsin, 445
            Henry Mall, Madison, WI 53706, USA
FEATURES             Location/Qualifiers
     source          1. .9280
                     /organism="Yersinia pestis KIM"
                     /strain="KIM"
                     /db_xref="taxon:187410"
     gene            complement(102. .1019)
                     /gene="gcvA"
                     /note="y3154"
     CDS             complement(102. .1019)
                     /gene="gcvA"
                     /function="regulator; central intermediary metabolism:
                     Pool, multipurpose conversions"
                     /note="residues 1 to 304 of 305 are 88.15 pct identical to
                     residues 1 to 304 of 305 from E. coli K12 : B2808;
                     residues 1 to 304 of 305 are 89.47 pct identical to
                     residues 1 to 304 of 305 from GenPept : >gb|AAL21860.1|
                     (AE008836) regulator of gcv operon (LysR family)
                     [Salmonella typhimurium LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="positive regulator of gcv operon"
                     /protein_id="AAM86704.1"
                     /db_xref="GI:21960079"
                     /translation="MSKRLPPLNALRAFDAAARHLSFTKAAEELFVTQAAVSHQIKSL
                     EDFLGLKLFRRRNRSLLLTEEGQSYYLDIKEIFTSINEATRKLQARSAKGALTVSLPP
                     SFAIQWLVPRLSGFNAAYPGIDVRIQAVDREEDKLADDVDVAIFYGRGNWSGLRTERL
                     YAEFLLPVCAPSLLTGENGLKVPSDLANHTLLHDTSRRDWLAYTRQLGVPQINVQQGP
                     IFSHSAMVVQAAVHGQGIALVNNVMAQSEIEAGRLVCPFNDVLVSKNAFYLVCHDSQA
                     ELGKIAAFRQWILARAASEQEKLRFRYEN"
     gene            1510. .1692
                     /gene="y3155"
     CDS             1510. .1692
                     /gene="y3155"
                     /function="unknown"
                     /note="residues 12 to 60 of 60 are 35.08 pct identical to
                     residues 559 to 613 of 752 from GenPept :
                     >gb|AAK76896.1|AE001438_149 (AE001438) Integrin-like
                     repeats domain fused to lysozyme, LYCV glycosyl hydrolase
                     [Clostridium acetobutylicum]"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical"
                     /protein_id="AAM86705.1"
                     /db_xref="GI:21960080"
                     /translation="MTDHLYLIVYFRFTISNDSATITNVYNVNMHNIIAQTRDHDKSK
                     KISGSLTGCQQQYESF"
     gene            1679. .2884
                     /gene="y3156"
     CDS             1679. .2884
                     /gene="y3156"
                     /note="residues 1 to 401 of 401 are 67.08 pct identical to
                     residues 1 to 401 of 401 from E. coli K12 : B2810"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM86706.1"
                     /db_xref="GI:21960081"
                     /translation="MKVFNPMDFRREFPALSDKLTYLDSAATALKPRAMIDATQQFYQ
                     QDSATVHRSQHQSALSLTVRFENTRQQVADFINSSTAENIIWTRGTTEAINLIAQSYA
                     RPRLQPEDEIIVSEAEHHANLIPWLMVAEQTGAKIVKLPLGLDHLPDLQQLPQLLNEK
                     TRILALGQMSNVTGGSPDLAQAIRLAHQYDCVVVVDGAQGIVHYPADVQALDIDFYAF
                     SSHKLYGPTGIGVLYGKTELLEEMPAWQGGGKMLTHASFGGFTPHEVPYRFEAGTPNI
                     AGVIGLSAVLKWLEHIDLEEAEVYSQGLATMAENKLAQLPGFHSYRCQQSSLLAFTFD
                     GVHHSDLVALLAEQGIALRAGQHCAQPLMAALGVNGSLRASFAPYNTPQDVEMLCSAL
                     GKALELLRD"
     gene            3019. .3462
                     /gene="y3157"
     CDS             3019. .3462
                     /gene="y3157"
                     /note="residues 4 to 137 of 147 are 58.95 pct identical to
                     residues 7 to 140 of 147 from E. coli K12 : B2811"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAM86707.1"
                     /db_xref="GI:21960082"
                     /translation="MMIAPHPFGHDITATDLIEKFSAHKQWEDRYRQLILLAKQLPPL
                     QEAWKKNELELTGCENRVWLGHQHLPDGTLHFYGDSEGRIVRGLLAVILTAVEGKTPQ
                     QVLAADPLALFEQLGLRQQLSTSRANGLQALAQGVQTIAAKYAER"
     gene            complement(3498. .4325)
                     /gene="y3158"
     CDS             complement(3498. .4325)
                     /gene="y3158"
                     /function="putative"
                     /note="residues 1 to 267 of 275 are 80.52 pct identical to
                     residues 1 to 267 of 268 from E. coli K12 : B2812;
                     residues 1 to 267 of 275 are 80.89 pct identical to
                     residues 1 to 267 of 268 from GenPept : >gb|AAL21865.1|
                     (AE008837) paral putative enzyme [Salmonella typhimurium
                     LT2]"
                     /codon_start=1
                     /transl_table=11
                     /product="putative enzyme"
                     /protein_id="AAM86708.1"
                     /db_xref="GI:21960083"
                     /translation="MSTAYSEAYQQRFGGIARLYGQQALALFSQAHVCVIGIGGVGSW
                     AAEALARTGIGAITLIDMDDVCVTNTNRQIHALRHNIGQAKTEVMAERILAINPECHV
                     TCIDDFITADNVAELLNKNFSYVIDAIDSVRPKAALLSYCRRYKIPVVTTGGAGGQID
                     PTRIAVVDLAKTIQDPLAAKLRERLKSDFNVVKNSKGKLGIDCVFSSEPLVYPQADGS
                     VCASRSTAEGPKKMDCTSGFGSATMVTATFGFVAVSHVLKKMMAKAARQHESAATAS"
     gene            complement(4468. .5640)
                     /gene="mltA"
                     /note="y3159"
     CDS             complement(4468. .5640)
                     /gene="mltA"
                     /function="enzyme; murein sacculus, peptidoglycan"
                     /note="residues 1 to 366 of 390 are 69.94 pct identical to
                     residues 1 to 364 of 365 from E. coli K12 : B2813"
                     /codon_start=1
                     /transl_table=11
                     /product="membrane-bound lytic murein transglycosylase A"
                     /protein_id="AAM86709.1"
                     /db_xref="GI:21960084"
                     /translation="MTSRWGKYLLSGIMIAVLAGCQSRPTDRGQQYKDGRLEQSLELV
                     NEPNAAGKPVNAKDYSDQVKVINQSSPGLYNRNSDTFNAVQNWMLAGADTSKLSLFGL
```

```
                NAYQMEGVDNFGNVQFTGYYTPVLQARYTPQGEFRHPLYRMPAKGKRRLPDRAAIYAG
                ALDNRNLIIAYTNSLVDNFMMEVQGSGYVDYGDGRPLTFFGYAGKNGHAYRSIGKVLI
                DRGEVARADMSMQAIRQWAENHSEAEVRELLEQNPSFVFFKPVMYAPVKGASAVPLIA
                KASVASDKSLIPPGTTLLAEVPLLDDQGKFTGKYQMRLMVALDVGGAIKGQHFDIYQG
                IGHEAGQAAGFYNHYGRVWVLKNAQSSGPLFTAYKGGTQSEPTSNDSSLLVNNQDR"
     gene       5520. .5702
                /gene="y3160"
     CDS        5520. .5702
                /gene="y3160"
                /function="unknown"
                /note="residues 6 to 41 of 60 are 38.88 pct identical to
                residues 143 to 178 of 327 from GenPept : >emb|CAA22272.1|
                (AL034381) conserved hypothetical protein
                [Schizosaccharomyces pombe]"
                /codon_start=1
                /transl_table=11
                /product="hypothetical"
                /protein_id="AAM86710.1"
                /db_xref="GI:21960085"
                /translation="MIAPGDHPYIVAHDRLAVIGNRPEQQSLFHSVGICPNDLSFHAF
                SLTHFIYRRTITNAYG"
     tRNA       5856. .5929
                /product="tRNA-Met"
                /note="anticodon: CAT"
                /function="RNA; tRNA"
     tRNA       6056. .6129
                /product="tRNA-Met"
                /note="anticodon: CAT"
                /function="RNA; tRNA"
     gene       complement(6362. .7630)
                /gene="y3161"
     CDS        complement(6362. .7630)
                /gene="y3161"
                /function="putative"
                /note="residues 1 to 421 of 422 are 69.88 pct identical to
                residues 25 to 445 of 447 from E. coli K12 : B2817;
                residues 7 to 421 of 422 are 72.31 pct identical to
                residues 1 to 415 of 417 from GenPept : >emb|CAD02815.1|
                (AL627277) N-acetylmuramoyl-L-alanine amidase [Salmonella
                enterica subsp. enterica serovar Typhi]"
                /codon_start=1
                /transl_table=11
                /product="putative amidase"
                /protein_id="AAM86711.1"
                /db_xref="GI:21960086"
                /translation="MLLWSYMADSNHNSGRRRLLQGAAAAWMLSISRVGFAASSHIVA
                VRVWPSSTYTRVTLESNTPLKYRQFALTHPDRIVVDIEGVQLNSVLKEISRQVQSADP
                YLKQARVGQFDKNTVRLVLELKQSISPQLFTLKPFAEFRNRLVVDLYPAEGGTSAEDD
                PLLALLEDYNKGNVDRTLPPETPKAGKAGRDRPIIIMLDPGHGGEDPGAIGKNKTKEK
                DIVLQIARRLRALIQKEANMRVFMTRNEDVFIPLKVRVAKARKLRADLFISIHADAFT
                NRAASGSSVFALSTKGATSTAARFLAQTQNEADQIGGVSKSGDPYLDHTIIDLLQTAT
                INDSLKFGKEVLSRMSKINKLHKNRVDQAGFAVLKAPDIPSILVETAFISNLEEERKL
                RTSRFQQQIAESIFAGIKAYFANGGAMARL"
     gene       7848. .9173
                /gene="argA"
                /note="y3162"
     CDS        7848. .9173
                /gene="argA"

  Query Match           42.7%;  Score 514.8;  DB 1;  Length 9280;
  Best Local Similarity  64.2%;  Pred. No. 1.8e-122;
  Matches  774;  Conservative    0;  Mismatches  432;  Indels    0;  Gaps    0;

Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          |||||| ||||||||||||   | | |||||     | || || || ||    |||
Db   1679 ATGAAAGTTTTTAATCCAATGGATTTTCGTCGGGAATTCCCTGCGCTCAGTGATAAATTA 1738

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          |||||| ||||| || ||||||||  | |||||    ||  || |||| || || ||
Db   1739 ACCTATCTGGACAGTGCGGCGACCGCCTTGAAACCACGTGCAATGATTGACGCGACACAG 1798

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          || ||||||    | |    | | ||  || || ||||||||   |      || |

Db   1799 CAATTTTATCAGCAGGATTCAGCAACGGTACACCGCAGCCAACATCAATCGGCGCTGTCA 1858

Qy    181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          | || |  || | ||||    | ||  |  |||||||  |||  | ||  || |
Db   1859 TTAACGGTTCGCTTTGAAAACACCCGCCAACAAGTGGCTGATTTTATTAACTCATCTACA 1918

Qy    241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          |   ||| |||  |||||||||| || || ||||||  |||||||  || | || ||| |
Db   1919 GCAGAAAATATTATCTGGACGCGAGGAACAACTGAAGCCATCAATCTGATCGCGCAAAGT 1978

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||| || || ||| | ||||| |  ||||| |||||||||||||  || ||||| ||
Db   1979 TATGCCCGCCCCCGTTTACAACCTGAAGATGAAATTATTGTCAGCGAAGCTGAACATCAT 2038

Qy    361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          || ||  |  | |||||| ||||||| ||  | || || || || ||| | || |||||
Db   2039 GCGAATTTAATTCCCTGGTTGATGGTAGCGGAGCAGACCGGTGCAAAAATAGTCAAATTA 2098

Qy    421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          || |||       | |  ||||| ||| |  |     | || ||||  ||| |
Db   2099 CCTCTTGGCCTTGATCATCTGCCAGATTTACAGCAACTCCCTCAACTACTTAATGAAAAA 2158

Qy    481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          |  || ||  | ||| |||| |||||||| ||||| || |||||| |||| |||||||
Db   2159 ACACGCATATTAGCGCTGGGGCAGATGTCTAACGTAACAGGCGGTAGCCCTGATCTGGCT 2218

Qy    541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          |  || ||||  | |||||  |  |    || || ||||||| ||||||||||
Db   2219 CAGGCTATTAGGCTGGCTCACCAATATGACTGTGTTGTCGTGGTTGACGGTGCTCAGGGA 2278

Qy    601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
            || |||| ||| || |||||||||  | ||||||||||||| ||||| || ||
Db   2279 ATTGTTCATTACCCAGCCGATGTTCAGGCATTGGATATTGATTTTTATGCATTCTCTTCC 2338

Qy    661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          || ||| |||||||||| || || || |||||||||||||| ||  | ||| |  | ||
Db   2339 CATAAATTGTATGGCCCAACCGGCATTGGCGTGCTGTATGGGAAGACTGAATTATTAGAA 2398

Qy    721 GCGATGTCGCCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          | |||| |  |||||| |||||||| |||||| ||  | | |     || |  |||||
Db   2399 GAGATGCCCGCCTGGCAAGGCGGCGGTAAAATGCTTACCCATGCATCATTCGGGGGCTTT 2458

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||  ||||    | ||| |       | ||||| || || || ||| | ||||| || ||
Db   2459 ACACCTCATGAAGTGCCTTATCGCTTTGAAGCGGGTACACCCAATATTGCTGGCGTTATT 2518

Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          || |||   |||| ||  |||||||| | ||   ||| |  |  |||||||    |
Db   2519 GGTTTATCAGCGGTACTCAAATGGCTGGAACATATTGATCTGGAAGAGGCCGAAGTTTAT 2578

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTGCCGGCTTTCGTTCA 960
          ||||   | ||||| ||  |||| ||| ||  ||| ||  ||  || || ||||
Db   2579 AGCCAAGGTTTAGCTACAATGGCAGAAAATAAGCTCGCACAATTACCGGGTTTTCACAGT 2638

Qy    961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          | ||| || ||| | |||||  |  | || |||  ||| | ||| |||||||| || |||
Db   2639 TACCGTTGTCAGCAATCCAGTTTATTAGCATTTACTTTCGATGGTGTTCATCACAGTGAT 2698

Qy   1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          | ||| || |  |||| ||| | ||||| || || || || |||||||| ||||| |||
Db   2699 TTAGTGGCGTTATTGGCCGAGCAAGGTATCGCACTACGTGCTGGGCAGCACTGCGCACAG 2758

Qy   1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          || ||  |||| |   | || || |  |||| |||| || ||||||||||| ||||||
Db   2759 CCACTGATGGCGGCTCTGGGAGTCAATGGCAGTCTGCGGGCTTCTTTTGCGCCTTATAAT 2818

Qy   1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCGCTGGAATTATTGGTG 1200
          ||       ||||| ||  |||       ||| |||     ||  ||||| |  |
Db   2819 ACCCCCCAAGATGTTGAAATGCTTTGCTCTGCGCTTGGTAAGGCATTGGAACTGCTTCGA 2878

Qy   1201 GATTAA 1206
          || |||
Db   2879 GACTAA 2884
```

Search completed: April 22, 2003, 12:23:01
Job time : 3767.45 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 08:16:32 ; Search time 302.905 Seconds
(without alignments)
8966.219 Million cell updates/sec

Title: US-09-622-419-3
Perfect score: 1206
Sequence: 1 atgaacgtttttaatcccgc.....tggaattattggtggattaa 1206

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1206	100.0	1206	20	AAZ09785	E. coli bioS1 DNA.
	2	1206	100.0	3465	20	AAX02814	DE19731274 Seq ID
	3	1206	100.0	3794	20	AAZ09791	Plasmid pHS1 bioS1
	4	1206	100.0	3794	20	AAX02812	DE19731274 Seq ID
	5	1206	100.0	4975	20	AAZ09792	Plasmid pHS1 metK/
	6	1205	99.9	1216	20	AAX02810	E. coli biotin ORF
	7	369	30.6	524	24	ABQ48374	Oligonucleotide fo
c	8	369	30.6	524	24	ABQ48375	Oligonucleotide fo
c	9	348	28.9	524	24	ABQ48376	Oligonucleotide fo
	10	348	28.9	524	24	ABQ48377	Oligonucleotide fo
	11	287.6	23.8	917	24	ABQ23332	Oligonucleotide fo
c	12	287.6	23.8	917	24	ABQ23333	Oligonucleotide fo
c	13	277	23.0	917	24	ABQ23330	Oligonucleotide fo
	14	277	23.0	917	24	ABQ23331	Oligonucleotide fo
	15	244	20.2	1221	20	AAZ09787	E. coli bioS3 DNA.
	16	234.6	19.5	1217	24	ABK77448	Bacillus clausii g
	17	234.6	19.5	1224	24	ABK77447	Bacillus clausii g
c	18	225	18.7	4160	24	ABQ71003	Listeria monocytog
c	19	215.4	17.9	684707	24	ABQ67196	Listeria innocua c
c	20	215.4	17.9	3011208	24	ABQ69245	Listeria innocua D
c	21	207.4	17.2	2944528	24	ABA03041	Listeria monocytog
c	22	204.8	17.0	1830121	17	AAT42063	Haemophilus influe
	23	198.4	16.5	10397	20	AAX13227	Enterococcus faeca
	24	192.4	16.0	18627	19	AAV52246	Streptococcus pneu
c	25	190.6	15.8	1035	24	ABQ68818	Listeria monocytog
c	26	175.2	14.5	4280	18	AAV74392	Staphylococcus aur
	27	170.2	14.1	598	23	AAS84162	DNA encoding novel
c	28	165.2	13.7	2365589	24	ABA90521	Genomic sequence o
	29	160	13.3	1224	24	ABN68933	Streptococcus poly
	30	153.8	12.8	1230	24	ABN68932	Streptococcus poly
	31	153.8	12.8	2155561	24	ABN71527	Streptococcus poly
c	32	153.4	12.7	1230025	20	AAX91990	Nucleotide sequenc
	33	151.2	12.5	1287	24	ABN92579	Staphylococcus epi
	34	148	12.3	4973	19	AAV58289	S. pyogenes SP-55-
	35	145.2	12.0	727	24	ABK73013	Bacillus lichenifo
c	36	138	11.4	265	23	AAS51198	Salmonella typhimu
c	37	129	10.7	13256	20	AAX20528	Polynucleotide seq
c	38	125.4	10.4	21719	23	AAS59568	Propionibacterium
	39	119.4	9.9	1263	22	AAH66692	C glutamicum codin
	40	119.4	9.9	1386	22	AAF72081	Corynebacterium gl
c	41	119.4	9.9	349980	22	AAH68529	C glutamicum codin
	42	99.2	8.2	534	24	ABN78987	Human isomerase-li
	43	90.4	7.5	761	22	AAF25823	C. glutamicum nifS
	44	87.6	7.3	349980	22	AAF86431	Pyrococcus abyssi
	45	74	6.1	4443	23	AAS71615	DNA encoding novel

ALIGNMENTS

RESULT 1
AAZ09785
ID AAZ09785 standard; DNA; 1206 BP.
XX
AC AAZ09785;
XX
DT 23-NOV-1999 (first entry)
XX
DE E. coli bioS1 DNA.
XX
KW SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW bioS3; cofactor; decarboxylation; Vitamin H; ss.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..1206
FT /*tag= a
FT /product= "bioS1"
XX
PN DE19806872-A1.
XX
PD 26-AUG-1999.
XX
PF 19-FEB-1998; 98DE-1006872.
XX
PR 19-FEB-1998; 98DE-1006872.
XX
PA (BADI) BASF AG.
XX
PI Schroeder H;
XX

```
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33264.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
XX
PS   Claim 1; Page 12-14; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microprganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the Escherichia coli bioS1 protein which is used in the method of the
CC   invention.
XX
SQ   Sequence 1206 BP; 256 A; 315 C; 359 G; 276 T; 0 other;

  Query Match             100.0%;  Score 1206;  DB 20;  Length 1206;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
Db       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
Db      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
Db     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
Db     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
Db     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
Db     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
Db     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
Db     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
Db     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
Db     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
Db     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
Db     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
Db     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
Db     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
Db     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
Db     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
Db     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
Db    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
Db    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
Db    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200

Qy    1201 GATTAA 1206
Db    1201 GATTAA 1206


RESULT 2
AAX02814
ID   AAX02814 standard; DNA; 3465 BP.
XX
AC   AAX02814;
XX
DT   14-MAY-1999  (first entry)
XX
DE   DE19731274 Seq ID 9.
XX
KW   Biotin; synthesis; dethiobiotin; yield increase; industry;
KW   fermentation; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             272..1477
FT                   /*tag= a
XX
PN   DE19731274-A1.
XX
PD   28-JAN-1999.
XX
PF   22-JUL-1997;   97DE-1031274.
XX
PR   22-JUL-1997;   97DE-1031274.
XX
PA   (BADI ) BASF AG.
XX
PI   Hauer B,  Schroeder H;
XX
DR   WPI; 1999-107030/10.
DR   P-PSDB; AAW92938.
XX
PT   Improved synthesis of biotin by expressing the bioS1 or bioS2
```

```
PT   sequence in biotin-producing cells - and related gene constructs,
PT   provides increased conversion of dethiobiotin in eukaryotic or
PT   prokaryotic hosts
XX
PS   Example 2; Page 32-35; 48pp; German.
XX
CC   This invention describes a method for the synthesis of biotin in
CC   Escherichia coli. This method involves the expression of a biotin gene
CC   or its functional variants, analogues and derivatives, in a prokaryote
CC   or eukaryote that is able to produce biotin. The cells are grown and
CC   the biotin produced either used directly, after separation of the
CC   biomass, or after purification. Constructs containing this nucleic acid
CC   or protein or their variants etc., can be coupled to one or more
CC   regulators for increasing gene and/or protein expression, and/or having
CC   its natural regulators 'switched off'. Expression of this biotin protein
CC   leads to increased conversion, by at least 3-fold, of dethiobiotin to
CC   biotin, thus increasing yield and making possible an industrially useful
CC   fermentative method for biotin production.
XX
SQ   Sequence 3465 BP; 804 A; 963 C; 934 G; 764 T; 0 other;

  Query Match             100.0%;  Score 1206;  DB 20;  Length 3465;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     272 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 331

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     332 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 391

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 451

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     452 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 511

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     512 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 571

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     572 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 631

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     632 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 691

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     692 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 751

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     752 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 811

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     812 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 871

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     872 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 931

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     932 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 991

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     992 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1051

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1052 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1111

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1112 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1171

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1172 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1231

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1232 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1291

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1292 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1351

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1352 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1411

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1412 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1471

Qy    1201 GATTAA 1206
          ||||||
Db    1472 GATTAA 1477


RESULT 3
AAZ09791
ID   AAZ09791 standard; DNA; 3794 BP.
XX
AC   AAZ09791;
XX
DT   23-NOV-1999  (first entry)
XX
DE   Plasmid pHS1 bioS1 DNA.
XX
KW   SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW   bioS3; cofactor; decarboxylation; Vitamin H; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             601..1806
FT                   /*tag= a
FT                   /product= "bioS1"
XX
PN   DE19806872-A1.
XX
PD   26-AUG-1999.
XX
PF   19-FEB-1998;   98DE-1006872.
XX
PR   19-FEB-1998;   98DE-1006872.
XX
PA   (BADI ) BASF AG.
XX
PI   Schroeder H;
XX
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33268.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
```

```
PT   second biotin synthesis gene in host cells -
XX
PS   Example 3; Page 30-33; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the bioS1 protein from plasmid pHS1 which is used in the method of the
CC   invention.
XX
SQ   Sequence 3794 BP; 910 A; 1034 C; 1020 G; 830 T; 0 other;

  Query Match             100.0%;  Score 1206;  DB 20;  Length 3794;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 660

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 720

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 780

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 840

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 900

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 960

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 1020

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 1080

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 1140

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 1200

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 1260

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 1320

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1380

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1440

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1500

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1560

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1620

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1680

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1740

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCGCTGGAATTATTGGTG 1800

Qy    1201 GATTAA 1206
          ||||||
Db    1801 GATTAA 1806


RESULT 4
AAX02812
ID   AAX02812 standard; DNA; 3794 BP.
XX
AC   AAX02812;
XX
DT   14-MAY-1999  (first entry)
XX
DE   DE19731274 Seq ID 5.
XX
KW   Biotin; synthesis; dethiobiotin; yield increase; industry;
KW   fermentation; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             601..1806
FT                   /*tag= a
XX
PN   DE19731274-A1.
XX
PD   28-JAN-1999.
XX
PF   22-JUL-1997;   97DE-1031274.
XX
PR   22-JUL-1997;   97DE-1031274.
XX
PA   (BADI ) BASF AG.
XX
PI   Hauer B,  Schroeder H;
XX
DR   WPI; 1999-107030/10.
DR   P-PSDB; AAW92936.
XX
PT   Improved synthesis of biotin by expressing the bioS1 or bioS2
PT   sequence in biotin-producing cells - and related gene constructs,
PT   provides increased conversion of dethiobiotin in eukaryotic or
PT   prokaryotic hosts
XX
```

```
PS   Example 2; Page 21-24; 48pp; German.
XX
CC   This invention describes a method for the synthesis of biotin in
CC   Escherichia coli. This method involves the expression of a biotin gene
CC   or its functional variants, analogues and derivatives, in a prokaryote
CC   or eukaryote that is able to produce biotin. The cells are grown and
CC   the biotin produced either used directly, after separation of the
CC   biomass, or after purification. Constructs containing this nucleic acid
CC   or protein or their variants etc., can be coupled to one or more
CC   regulators for increasing gene and/or protein expression, and/or having
CC   its natural regulators 'switched off'. Expression of this biotin protein
CC   leads to increased conversion, by at least 3-fold, of dethiobiotin to
CC   biotin, thus increasing yield and making possible an industrially useful
CC   fermentative method for biotin production.
XX
SQ   Sequence 3794 BP; 910 A; 1034 C; 1020 G; 830 T; 0 other;

  Query Match             100.0%;  Score 1206;  DB 20;  Length 3794;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 660

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 720

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 780

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 840

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 900

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 960

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 1020

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 1080

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 1140

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 1200

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 1260

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 1320

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1380

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1440

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1500

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1560

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1620

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1680

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1740

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1800

Qy    1201 GATTAA 1206
          ||||||
Db    1801 GATTAA 1806


RESULT 5
AAZ09792
ID   AAZ09792 standard; DNA; 4975 BP.
XX
AC   AAZ09792;
XX
DT   23-NOV-1999  (first entry)
XX
DE   Plasmid pHS1 metK/bioS1 fusion protein DNA.
XX
KW   SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW   bioS3; cofactor; decarboxylation; Vitamin H; metK; ss.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   CDS             530..1684
FT                   /*tag= a
FT                   /product= "metK"
FT   CDS             1782..2987
FT                   /*tag= b
FT                   /product= "bioS1"
XX
PN   DE19806872-A1.
XX
PD   26-AUG-1999.
XX
PF   19-FEB-1998;   98DE-1006872.
XX
PR   19-FEB-1998;   98DE-1006872.
XX
PA   (BADI ) BASF AG.
XX
PI   Schroeder H;
XX
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33269, AAY33270.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
```

```
XX
PS   Example 3; Page 36-41; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the metK/bioS1 fusion protein found in plasmid pHS1 which is used in the
CC   method of the invention.
XX
SQ   Sequence 4975 BP; 1198 A; 1390 C; 1329 G; 1098 T; 0 other;

  Query Match           100.0%;  Score 1206;  DB 20;  Length 4975;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1782 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 1841

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1842 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 1901

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1902 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 1961

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1962 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 2021

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2022 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 2081

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2082 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 2141

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2142 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 2201

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2202 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 2261

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2262 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 2321

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2322 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 2381

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2382 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 2441

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2442 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 2501

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2502 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 2561

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2562 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 2621

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2622 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 2681

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2682 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 2741

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2742 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 2801

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2802 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 2861

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2862 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 2921

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2922 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 2981

Qy    1201 GATTAA 1206
          ||||||
Db    2982 GATTAA 2987


RESULT 6
AAX02810
ID   AAX02810 standard; DNA; 1216 BP.
XX
AC   AAX02810;
XX
DT   14-MAY-1999  (first entry)
XX
DE   E. coli biotin ORF401 DNA.
XX
KW   Biotin; ORF401; synthesis; dethiobiotin; yield increase; industry;
KW   fermentation; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             12..1216
FT                   /*tag= a
FT                   /product= "ORF401"
XX
PN   DE19731274-A1.
XX
PD   28-JAN-1999.
XX
PF   22-JUL-1997;   97DE-1031274.
XX
PR   22-JUL-1997;   97DE-1031274.
XX
PA   (BADI ) BASF AG.
XX
PI   Hauer B,  Schroeder H;
XX
DR   WPI; 1999-107030/10.
DR   P-PSDB; AAW92934.
XX
PT   Improved synthesis of biotin by expressing the bioS1 or bioS2
PT   sequence in biotin-producing cells - and related gene constructs,
PT   provides increased conversion of dethiobiotin in eukaryotic or
PT   prokaryotic hosts
XX
```

PS Claim 1; Page 12-14; 48pp; German.
XX
CC This invention describes a method for the synthesis of biotin in
CC Escherichia coli. This method involves the expression of a biotin gene
CC or its functional variants, analogues and derivatives, in a prokaryote
CC or eukaryote that is able to produce biotin. The cells are grown and
CC the biotin produced either used directly, after separation of the
CC biomass, or after purification. Constructs containing this nucleic acid
CC or protein or their variants etc., can be coupled to one or more
CC regulators for increasing gene and/or protein expression, and/or having
CC its natural regulators 'switched off'. Expression of this biotin protein
CC leads to increased conversion, by at least 3-fold, of dethiobiotin to
CC biotin, thus increasing yield and making possible an industrially useful
CC fermentative method for biotin production.
XX
SQ Sequence 1216 BP; 258 A; 318 C; 363 G; 277 T; 0 other;

Query Match 99.9%; Score 1205; DB 20; Length 1216;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      12 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 71

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      72 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 131

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     132 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 191

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     192 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 251

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     252 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 311

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     312 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 371

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     372 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 431

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 491

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     492 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 551

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     552 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 611

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     612 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 671

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     672 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 731

Qy     721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     732 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 791

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     792 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 851

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     852 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 911

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     912 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 971

Qy     961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     972 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1031

Qy    1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1032 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1091

Qy    1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1092 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1151

Qy    1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1152 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1211

Qy    1201 GATTA 1205
          |||||
Db    1212 GATTA 1216
```

RESULT 7
ABQ48374
ID ABQ48374 standard; DNA; 524 BP.
XX
AC ABQ48374;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34965.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 524 BP; 114 A; 44 C; 152 G; 214 T; 0 other;

Query Match 30.6%; Score 369; DB 24; Length 524;
Best Local Similarity 81.8%; Pred. No. 3.1e-106;
Matches 426; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```
Qy    119 AACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAAC 178
          || ||||||| ||| |||||| |||||||| |||| ||  |||||| |||||   |||
Db      4 AATAGTTTTATAGTTTGAGCGTCGGAAACGTTTATCGTAGTTAGTTTGTCGAAGTTTAAC 63

Qy    179 GCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGG 238
          |  ||| |||||||||||||| || |||||||||||||||| ||||| |||||| | |||
Db     64 GTTTGATCGCGCGTTATGAAGTTGTACGAGAGAAAGTGGCGTAATTATTGAATGTATCGG 123

Qy    239 ATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAAT 298
          ||||| ||| |||||| |||||||||||  |  | |||||  || || ||||||| | |||
Db    124 ATGATTAAATTATCGTTTGGACGCGCGGTATTATTGAATTTATTAATATGGTGGTATAAT 183

Qy    299 GCTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACC 358
          | ||||||||| ||||| || || |||||||||||||||||| ||||||| |||| |
Db    184 GTTATGCGCGTTCGCGTTTGTAATCGGGCGATGAGATTATTGTTAGCGTGGTAGAATATT 243

Qy    359 ACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAAT 418
          |||  ||  ||||    ||| |||||||||   || ||| |||||  ||||||||||||
Db    244 ACGTTAATTTCGTTTTTTGGTTGATGGTCGTTTAATAAATTGGAGTTAAAGTGGTGAAAT 303

Qy    419 TGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCC 478
          || || |||||||| ||||| || ||||||||||||||||  |||| ||||||  |   |
Db    304 TGTCGTTTAATGCGTAGCGATTGTCGGATGTCGATTTGTTGTTAGAATTGATTATTTTTC 363

Qy    479 GTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGG 538
          |||||||||||  |||||||||| |||||||||||||| |||||||||  ||||| |||
Db    364 GTAGTCGGATTTTGGCGTTGGGTTAGATGTCGAACGTTATTGGCGGTTGTTCGGATTTGG 423

Qy    539 CGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGG 598
          ||||||||||||  |||| | ||| || |||||||||||||||||||||||||||| | |||
Db    424 CGCGAGCGATTATTTTTGTTTATTTAGTCGGGATGGTGGTGATGGTTGATGGTGTTTAGG 483

Qy    599 GGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATT 639
          ||| |||| ||||   |||||||||| || || ||||||||
Db    484 GGGTAGTGTATTTTTTCGCGGATGTTTAGTAATTGGATATT 524
```

RESULT 8
ABQ48375/c
ID ABQ48375 standard; DNA; 524 BP.
XX
AC ABQ48375;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34966.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 524 BP; 214 A; 152 C; 44 G; 114 T; 0 other;

Query Match 30.6%; Score 369; DB 24; Length 524;
Best Local Similarity 81.8%; Pred. No. 3.1e-106;
Matches 426; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

```
Qy    119 AACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAAC 178
          || ||||||| ||| |||||| |||||||| |||| ||  |||||| |||||   |||
Db    521 AATAGTTTTATAGTTTGAGCGTCGGAAACGTTTATCGTAGTTAGTTTGTCGAAGTTTAAC 462

Qy    179 GCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGG 238
          |  ||| |||||||||||||| || |||||||||||||||| ||||| |||||| | |||
Db    461 GTTTGATCGCGCGTTATGAAGTTGTACGAGAGAAAGTGGCGTAATTATTGAATGTATCGG 402

Qy    239 ATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAAT 298
          ||||| ||| |||||| |||||||||||  |  | |||||  || || ||||||| | |||
Db    401 ATGATTAAATTATCGTTTGGACGCGCGGTATTATTGAATTTATTAATATGGTGGTATAAT 342

Qy    299 GCTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACC 358
          | ||||||||| ||||| || || |||||||||||||||||| ||||||| |||| |
Db    341 GTTATGCGCGTTCGCGTTTGTAATCGGGCGATGAGATTATTGTTAGCGTGGTAGAATATT 282

Qy    359 ACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAAT 418
          |||  ||  ||||    ||| |||||||||   || ||| |||||  ||||||||||||
Db    281 ACGTTAATTTCGTTTTTTGGTTGATGGTCGTTTAATAAATTGGAGTTAAAGTGGTGAAAT 222
```

```
Qy    419 TGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCC 478
          || || |||||||| ||||| || ||||||||||||||||  |||| ||||||  |   |
Db    221 TGTCGTTTAATGCGTAGCGATTGTCGGATGTCGATTTGTTGTTAGAATTGATTATTTTTC 162

Qy    479 GTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGG 538
          |||||||||| |||||||||| |||||||||||||| |||||||||  ||||| |||
Db    161 GTAGTCGGATTTTGGCGTTGGGTTAGATGTCGAACGTTATTGGCGGTTGTTCGGATTTGG 102

Qy    539 CGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGG 598
          ||||||||||||  |||| | ||| || ||||||||||||||||||||||||| | |||
Db    101 CGCGAGCGATTATTTTTGTTTATTTAGTCGGGATGGTGGTGATGGTTGATGGTGTTTAGG 42

Qy    599 GGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATT 639
          ||| |||| ||||   |||||||||| || || ||||||||
Db     41 GGGTAGTGTATTTTTTCGCGGATGTTTAGTAATTGGATATT 1
```

RESULT 9
ABQ48376/c
ID ABQ48376 standard; DNA; 524 BP.
XX
AC ABQ48376;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34967.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA -
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 524 BP; 116 A; 44 C; 142 G; 222 T; 0 other;

Query Match 28.9%; Score 348; DB 24; Length 524;
Best Local Similarity 79.0%; Pred. No. 1.4e-99;
Matches 414; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```
Qy    116 CCCAACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCC 175
          ||||||| ||||||| ||| | ||||| ||||||||||||||| ||| ||| ||||| |||
Db    524 CCCAACAATTTTACAATCTAAACGCCGAAAACGTCCATCGCAACCAATTTACCGAAACCC 465

Qy    176 AACGCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCAC 235
          ||||||| |||||||||||| || || ||||| | ||| |  |||||||||| ||| |||
Db    464 AACGCCTAACCGCGCGTTATAAAACTACACGAAAAAAAATAACGCAATTACTAAATACAC 405

Qy    236 CGGATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCAC 295
          || || |  |||||||||||  ||||||| ||||||| ||||||||||||  |  |||
Db    404 CGAATAACTAAACTATCGTCTAAACGCGCGACACCACTAAATCCATCAACATAATAACAC 345

Qy    296 AATGCTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAAC 355
          ||| |||| |||||||||||||| ||||||  |||| | ||||||  ||| |||  || |||
Db    344 AATACTATACGCGTCCGCGTCTACAACCGAACGATAAAATTATTATCAACGTAACAAAAC 285

Qy    356 ACCACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGA 415
          ||||||||||||||||||||||  || ||  |||||||||||||  | ||||| |  | |
Db    284 ACCACGCCAACCTCGTCCCCTAACTAATAATCGCCCAACAAACTAAAACCAAAATAATAA 225

Qy    416 AATTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTC 475
          |||| |||||||| |||| ||||| ||| || ||||||| || ||| |||| |||||||
Db    224 AATTACCGCTTAATACGCAACGACTACCGAATATCGATTTATTACCAAAACTAATTACTC 165

Qy    476 CCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATC 535
          |||||| ||| ||||| |||| ||| || |||||||||||  || || |||| |||
Db    164 CCCGTAATCGAATTCTAACGTTAAATCAAATATCGAACGTTACTAACGATTACCCGAATC 105

Qy    536 TGGCGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTC 595
          |  ||||| |||||||||| |||||||| |||  ||  |  | ||  || ||  | |||
Db    104 TAACGCGAACGATTACCTTTACTCATTCAACCGAAATAATAATAATAATTAATAATACTC 45

Qy    596 AGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATT 639
          |     || | |||||||||||| || |||| |||||  |||||
Db     44 AAAAAACAATACATTTCCCCGCGAATATTCAACAACTAAATATT 1
```

RESULT 10
ABQ48377
ID ABQ48377 standard; DNA; 524 BP.
XX
AC ABQ48377;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 34968.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX

```
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 524 BP; 222 A; 142 C; 44 G; 116 T; 0 other;

  Query Match             28.9%;  Score 348;  DB 24;  Length 524;
  Best Local Similarity   79.0%;  Pred. No. 1.4e-99;
  Matches  414;  Conservative    0;  Mismatches  110;  Indels    0;  Gaps    0;

Qy     116 CCCAACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCC 175
           ||||||| ||||||| ||| | ||||| |||||||||||||| ||| ||| ||||| |||
Db       1 CCCAACAATTTTACAATCTAAACGCCGAAAACGTCCATCGCAACCAATTTACCGAAACCC 60

Qy     176 AACGCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCAC 235
           ||||||| |||||||||||| || || ||||| | ||| |  |||||||||| ||| |||
Db      61 AACGCCTAACCGCGCGTTATAAAACTACACGAAAAAAAATAACGCAATTACTAAATACAC 120

Qy     236 CGGATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCAC 295
           || || |  |||||||||||  ||||||| ||||||| |||||||||||||  |  |||
Db     121 CGAATAACTAAACTATCGTCTAAACGCGCGACACCACTAAATCCATCAACATAATAACAC 180

Qy     296 AATGCTATGCCGCTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAAC 355
           ||| |||| |||||||||||||| ||||||  |||| | |||||| ||| |||  || |||
Db     181 AATACTATACGCGTCCGCGTCTACAACCGAACGATAAAATTATTATCAACGTAACAAAAC 240

Qy     356 ACCACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGA 415
           |||||||||||||||||||||  || ||  ||||||||||||||  | ||||| |  | |
Db     241 ACCACGCCAACCTCGTCCCCTAACTAATAATCGCCCAACAAACTAAAACCAAAATAATAA 300

Qy     416 AATTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTC 475
           |||| |||||||||  |||| ||||| ||| || |||||||  || ||| |||| |||||||
Db     301 AATTACCGCTTAATACGCAACGACTACCGAATATCGATTTATTACCAAAACTAATTACTC 360

Qy     476 CCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCCGGATC 535
           |||||| ||| |||||  ||||   ||| || ||||||||||||  || || |||| |||
Db     361 CCCGTAATCGAATTCTAACGTTAAATCAAATATCGAACGTTACTAACGATTACCCGAATC 420

Qy     536 TGGCGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTC 595
           |  ||||| |||||||||||| |||||||| |||  ||  |  | ||  || ||  | |||
Db     421 TAACGCGAACGATTACCTTTACTCATTCAACCGAAATAATAATAATAATTAATAATACTC 480

Qy     596 AGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATT 639
           |     || | |||||||||||| || |||| |||||  |||||
Db     481 AAAAAACAATACATTTCCCCGCGAATATTCAACAACTAAATATT 524


RESULT 11
ABQ23332
ID   ABQ23332 standard; DNA; 917 BP.
XX
AC   ABQ23332;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 9923.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 917 BP; 201 A; 85 C; 250 G; 381 T; 0 other;

  Query Match             23.8%;  Score 287.6;  DB 24;  Length 917;
  Best Local Similarity   79.1%;  Pred. No. 2.4e-80;
  Matches  478;  Conservative    0;  Mismatches  114;  Indels   12;  Gaps   11;

Qy       9 TTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGCGTCTATCT 68
           |||||||  |||| |||||||||   |||||  || | || |||||||||||||| ||| |
Db      71 TTTTAATTTCGCGTAGTTTCGCGTTTAGTTTTTCGTATTATAGGATGCGGGCGT-TATTT 129

Qy      69 CGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAACAGTTTTA 128
```

```
Db    130 CGATAGCGTCGCGATCGCGTTTAAATTTGAAGTCGTGGTTGAA-TTATTTAATAGTTTTA 188

Qy    129 CAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGCCTGACCGC 188

Db    189 TAGTTTGAGCGTCGGAAACGTTTATCGTAGTTA-TTTGTCGAAGTTTAACGTTTGATCGC 247

Qy    189 GCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGATAAAAC 248

Db    248 GCGTTATGAAGTTGTACGAGAG-AAGTGGCGTAATTATTGAATGTATCGGATGATAAAAT 306

Qy    249 TATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCG 308

Db    307 TATCGTTTGGACG-GCGGTATTATTGAATTTATTAATATGGTGGTATAATGTTATGCGCG 365

Qy    309 TCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCT 368

Db    366 T-TTCGTTTGTAATCGGGCGATGAGATTATTGTTAGCGTGGTAGAATATTACGTAATTTC 424

Qy    369 CGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAA 428

Db    425 GTT--TTTGGTTGATGGTCGTTTAATAAATTGGAGTTAAAGTGG-GAAATTGTCGTTTAA 481

Qy    429 TGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGAT 488

Db    482 TGCGTAGCGATTGTCGGATGTCGATTTGTTGTTA-AATTGATTATTTTTCGTAGTCGGAT 540

Qy    489 TCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGAT 548

Db    541 TTTGGCGTTGGGTTAGATGTCGAA-GTTATTGGCGGTTGTTCGGATTTGGCGCGAGCGAT 599

Qy    549 TACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCA 608

Db    600 TATTTTTGTTTATT-AGTCGGGATGGTGGTGATGGTTGATGGTGTTTAGGGGGTAGTGTA 658

Qy    609 TTTC 612

Db    659 TTTC 662
```

RESULT 12
ABQ23333/c
ID ABQ23333 standard; DNA; 917 BP.
XX
AC ABQ23333;
XX
DT 12-JUL-2002 (first entry)
XX
DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 9924.
XX
KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW drug; side effect; cancer; central nervous system; cardiovascular;
KW gastrointestinal; respiratory system; single nucleotide polymorphism;
KW SNP; cell differentiation; ds.
XX
OS Homo sapiens.
XX
PN WO200218632-A2.
XX
PD 07-MAR-2002.
XX
PF 01-SEP-2001; 2001WO-EP10074.
XX
PR 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR WPI; 2002-371829/40.
XX
PT Determining the degree of cytosine methylation in genomic DNA, useful
PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.
XX
SQ Sequence 917 BP; 381 A; 250 C; 85 G; 201 T; 0 other;

Query Match 23.8%; Score 287.6; DB 24; Length 917;
Best Local Similarity 79.1%; Pred. No. 2.4e-80;
Matches 478; Conservative 0; Mismatches 114; Indels 12; Gaps 11;

```
Qy      9 TTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGCGTCTATCT 68

Db    847 TTTTAATTTCGCGTAGTTTCGCGTTTAGTTTTTCGTATTATAGGATGCGGGCGT-TATTT 789

Qy     69 CGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAACAGTTTTA 128

Db    788 CGATAGCGTCGCGATCGCGTTTAAATTTGAAGTCGTGGTTGAA-TTATTTAATAGTTTTA 730

Qy    129 CAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGCCTGACCGC 188

Db    729 TAGTTTGAGCGTCGGAAACGTTTATCGTAGTTA-TTTGTCGAAGTTTAACGTTTGATCGC 671

Qy    189 GCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGATAAAAC 248

Db    670 GCGTTATGAAGTTGTACGAGAG-AAGTGGCGTAATTATTGAATGTATCGGATGATAAAAT 612

Qy    249 TATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCG 308

Db    611 TATCGTTTGGACG-GCGGTATTATTGAATTTATTAATATGGTGGTATAATGTTATGCGCG 553

Qy    309 TCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCT 368

Db    552 T-TTCGTTTGTAATCGGGCGATGAGATTATTGTTAGCGTGGTAGAATATTACGTAATTTC 494

Qy    369 CGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAA 428

Db    493 GTT--TTTGGTTGATGGTCGTTTAATAAATTGGAGTTAAAGTGG-GAAATTGTCGTTTAA 437

Qy    429 TGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGAT 488

Db    436 TGCGTAGCGATTGTCGGATGTCGATTTGTTGTTA-AATTGATTATTTTTCGTAGTCGGAT 378

Qy    489 TCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGAT 548

Db    377 TTTGGCGTTGGGTTAGATGTCGAA-GTTATTGGCGGTTGTTCGGATTTGGCGCGAGCGAT 319

Qy    549 TACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCA 608

Db    318 TATTTTTGTTTATT-AGTCGGGATGGTGGTGATGGTTGATGGTGTTTAGGGGGTAGTGTA 260

Qy    609 TTTC 612
```

```
          ||||
Db    259 TTTC 256


RESULT 13
ABQ23330/c
ID   ABQ23330 standard; DNA; 917 BP.
XX
AC   ABQ23330;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 9921.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA  -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 917 BP; 202 A; 85 C; 264 G; 366 T; 0 other;

  Query Match             23.0%;  Score 277;  DB 24;  Length 917;
  Best Local Similarity   77.6%;  Pred. No. 5.5e-77;
  Matches  474;  Conservative    0;  Mismatches  125;  Indels   12;  Gaps   11;

Qy       4 AACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGCGTC 63
           ||   ||||||||||||||| |||||||||| |||||||||||||  || ||  || |
Db     852 AAACGTTTTAATCCCGCGCAATTTCGCGCCCAATTTCCCGCACTACAAAATACGAACG-C 794

Qy      64 TATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAACAG 123
           |||||||||| |||||||||||||||||||||||| || ||||  || || ||||||||||
Db     793 TATCTCGACAACGCCGCGACCGCGCTTAAACCTAAAACCGTAATT-AAACCACCCAACAA 735

Qy     124 TTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGCCTG 183
           ||||||| ||| | ||||| |||||||||||||| ||| ||| ||||| ||||||||||
Db     734 TTTTACAATCTAAACGCCGAAAACGTCCATCGCAACCA-TTTACCGAAACCCAACGCCTA 676

Qy     184 ACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGAT 243
           |||||||||||| || || ||||| | ||| |  |||||||||| ||| ||||| || ||
Db     675 ACCGCGCGTTATAAAACTACACGA-AAAAAATAACGCAATTACTAAATACACCGAATAAT 617

Qy     244 AAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTAT 303
           |||||||||||||  |||  || ||||||| |||||||||||||  |  |||||| ||||
Db     616 AAAACTATCGTCTAAACG-ACGACACCACTAAATCCATCAACATAATAACACAATACTAT 558

Qy     304 GCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCC 363
           |||||||  |||||  ||||||   |||| | ||||||  ||| |||  || ||||||||||
Db     557 ACGCGTCC-CGTCTACAACCGAACGATAAAATTATTATCAACGTAACAAAACACCACGC- 500

Qy     364 AACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCG 423
           | |     ||||  || ||  ||||||||||||||  | |||||  |    ||||| |||
Db     499 -AACCTCGTCCCTAACTAATAATCGCCCAACAAACTAAAACCAAAATAAA-AAATTACCG 442

Qy     424 CTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGT 483
           |||||| |||| ||||| ||| || ||||||| || ||| |||| |||||||||||| |
Db     441 CTTAATACGCAACGACTACCGAATATCGATTTATTACCA-AACTAATTACTCCCCGTAAT 383

Qy     484 CGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGA 543
           || |||||  ||||   ||| || |||||  |||||  || || |||| ||||  |||||
Db     382 CGAATTCTAACGTTAAATCAAATATCGAA-ATTACTAACGATTACCCGAATCTAACGCGA 324

Qy     544 GCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCA 603
           ||||||||||| |||||| | |||  ||  |  | ||  || ||  | ||||     ||
Db     323 ACGATTACCTTTACTCATT-AACCGAAATAATAATAATAATTAATAATACTCAAAAAACA 265

Qy     604 GTGCATTTCCC 614
           | |||||| |
Db     264 ATACATTTCGC 254


RESULT 14
ABQ23331
ID   ABQ23331 standard; DNA; 917 BP.
XX
AC   ABQ23331;
XX
DT   12-JUL-2002  (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 9922.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A,  Piepenbrock C,  Berlin K,  Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
```

```
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 917 BP; 366 A; 264 C; 85 G; 202 T; 0 other;

  Query Match             23.0%;  Score 277;  DB 24;  Length 917;
  Best Local Similarity   77.6%;  Pred. No. 5.5e-77;
  Matches  474;  Conservative    0;  Mismatches  125;  Indels   12;  Gaps   11;

Qy       4 AACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGCGTC 63
           ||   ||||||||||||||| |||||||||| ||||||||||||  || ||  || |
Db      66 AAACGTTTTAATCCCGCGCAATTTCGCGCCCAATTTCCCGCACTACAAAATACGAACG-C 124

Qy      64 TATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAACAG 123
           |||||||||| |||||||||||||||||||||| || ||||  || || ||||||||||
Db     125 TATCTCGACAACGCCGCGACCGCGCTTAAACCTAAAACCGTAATT-AAACCACCCAACAA 183

Qy     124 TTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGCCTG 183
           ||||||| ||| | ||||| |||||||||||||| ||| ||| ||||| ||||||||||
Db     184 TTTTACAATCTAAACGCCGAAAACGTCCATCGCAACCA-TTTACCGAAACCCAACGCCTA 242

Qy     184 ACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGAT 243
           |||||||||||| || || ||||| | ||| |  |||||||||| ||| ||||| || ||
Db     243 ACCGCGCGTTATAAAACTACACGA-AAAAAATAACGCAATTACTAAATACACCGAATAAT 301

Qy     244 AAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTAT 303
           ||||||||||||  |||  || ||||||| ||||||||||||  |  |||||| ||||
Db     302 AAAACTATCGTCTAAACG-ACGACACCACTAAATCCATCAACATAATAACACAATACTAT 360

Qy     304 GCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCC 363
           |||||||| ||||| ||||||  |||| | |||||| ||| |||  || ||||||||||
Db     361 ACGCGTCC-CGTCTACAACCGAACGATAAAATTATTATCAACGTAACAAAACACCACGC- 418

Qy     364 AACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCG 423
           | |    ||||  || ||  |||||||||||||| | ||||| |    ||||| |||
Db     419 -AACCTCGTCCCTAACTAATAATCGCCCAACAAACTAAAACCAAAATAAA-AAATTACCG 476

Qy     424 CTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGT 483
           |||||| |||| ||||| ||| || ||||||| || ||| |||| ||||||||||||| |
Db     477 CTTAATACGCAACGACTACCGAATATCGATTTATTACCA-AACTAATTACTCCCCGTAAT 535

Qy     484 CGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGA 543
           || |||||  ||||  ||| || |||||  |||||  || || |||| ||||  |||||
Db     536 CGAATTCTAACGTTAAATCAAATATCGAA-ATTACTAACGATTACCCGAATCTAACGCGA 594

Qy     544 GCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCA 603
           |||||||||||| |||||| | |||  ||  |  | ||  || ||  | ||||     ||
Db     595 ACGATTACCTTTACTCATT-AACCGAAATAATAATAATAATTAATAATACTCAAAAAACA 653

Qy     604 GTGCATTTCCC 614
           | ||||||  |
Db     654 ATACATTTCGC 664


RESULT 15
AAZ09787
ID   AAZ09787 standard; DNA; 1221 BP.
XX
AC   AAZ09787;
XX
DT   23-NOV-1999  (first entry)
XX
DE   E. coli bioS3 DNA.
XX
KW   SAM; S-adenosyl methionine synthase; biosynthesis; biotin; bioS1; bioS2;
KW   bioS3; cofactor; decarboxylation; Vitamin H; ss.
XX
OS   Escherichia coli.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..1221
FT                   /*tag= a
FT                   /product= "bioS3"
XX
PN   DE19806872-A1.
XX
PD   26-AUG-1999.
XX
PF   19-FEB-1998;   98DE-1006872.
XX
PR   19-FEB-1998;   98DE-1006872.
XX
PA   (BADI ) BASF AG.
XX
PI   Schroeder H;
XX
DR   WPI; 1999-480095/41.
DR   P-PSDB; AAY33266.
XX
PT   Production of biotin by expressing S-adenosyl-methionine synthase and
PT   second biotin synthesis gene in host cells -
XX
PS   Claim 1; Page 21-23; 48pp; German.
XX
CC   This invention describes a novel method for the preparation of biotin
CC   (I) which comprises expressing, in a prokaryotic or eukaryotic host
CC   capable of producing (I): (a) an S-adenosyl-methionine synthase (SAM)
CC   sequence (1), and (b) at least one of the other biotin biosynthesis
CC   genes bioS1, 2 or 3. (I) is a cofactor for enzyme-catalyzed
CC   (de)carboxylation reactions and is an essential vitamin (Vitamin H) for
CC   most animals and some microorganisms. Expression of biotin plus bioS1,
CC   bioS2 or bioS3 significantly increases productivity of biotin
CC   biosynthesis, particularly by at least 3 times. This sequence encodes
CC   the Escherichia coli bioS3 protein which is used in the method of the
CC   invention.
XX
SQ   Sequence 1221 BP; 264 A; 313 C; 357 G; 287 T; 0 other;

  Query Match             20.2%;  Score 244;  DB 20;  Length 1221;
  Best Local Similarity   52.3%;  Pred. No. 1.8e-66;
  Matches  588;  Conservative    0;  Mismatches  530;  Indels    6;  Gaps    2;

Qy      64 TATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAACAG 123
           |||||||||||||||||| |  ||||  |||||       ||| |||| |||       ||
Db      76 TATCTCGACAGCGCCGCCAGTGCGCAGAAACCGAGCCAGGTGATTGACGCCGAGGCCGAG 135

Qy     124 TTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGCCTG 183
           |||||  |||   ||  ||      || |||||  |      | ||  |  |  |    |
Db     136 TTTTATCGTCATGGCTACGCGGCGGTGCATCGTGGTATTCATACCTTAAGCGCCCAGGCG 195

Qy     184 ACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGAT 243
           ||||  |      ||    |  ||  ||   |  |||  ||  | |||||  |     |
```

```
Db    196 ACCGAGAAAATGGAGAACGTGCGCAAGCGGGCATCGCTGTTTATTAATGCCCGTTCGGCG 255

Qy    244 AAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTAT 303

Db    256 GAAGAGCTGGTGTTCGTCCGCGGCACGACGGAAGGGATCAATCTGGTCGCCAATAGCTGG 315

Qy    304 GCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCC 363

Db    316 GGCAACAGCAACGTGCGGGCGGGCGATAACATCATCATCAGTCAGATGGAGCACCACGCT 375

Qy    364 AACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCG 423

Db    376 AACATTGTTCCCTGGCAGATGCTTTGCGCACGCGTTGGCGCAGAGCTGCGTGTGATCCCG 435

Qy    424 CTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGT 483

Db    436 CTCAATCCCGATGGTACGTTGCAACTGGAGACGCTGCCTACGCTGTTTGATGAGAAAACT 495

Qy    484 CGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGA 543

Db    496 CGCCTGCTGGCAATTACTCATGTCTCCAACGTGCTTGGCACAGAAAATCCACTGGCGGAA 555

Qy    544 GCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCA 603

Db    556 ATGATCACGCTTGCGCACCAGCATGGCGCAAAAGTGCTGGTGGATGGCGCTCAGGCGGTG 615

Qy    604 GTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCAC 663

Db    616 ATGCATCATCCGGTGGATGTTCAGGCGCTGGATTGCGACTTTTACGTGTTCTCCGGGCAT 675

Qy    664 AAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCG 723

Db    676 AAACTGTATGGCCCCACCGGAATTGGCATTCTTTATGTGAAAGAAGCCTTGTTGCAGGAG 735

Qy    724 ATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTT---TGACGGCTTC 780

Db    736 ATGCCGCCGTGGGAAGGGGGCGGTTCTATGATCGCCACCGTCAGCCTGAGTGAAGGCACT 795

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840

Db    796 ACCTGGACCAAAGCACCATGGCGGTTTGAAGCCGGTACACCCAATACCGGGGGCATCATT 855

Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900

Db    856 GGTCTTGGCGCGGCGCTGGAGTATGTTTCGGCGCTGGGGCTTAATAACATAGCCGAGTAT 915

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960

Db    916 GAACAGAATCTGATGCATTATGCGCTATCACAGCTGGAATCTGTACCGGATCTCACTCTC 975

Qy    961 TTCCGCTGCCAGGA---TTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGC 1017

Db    976 TATGGCCCACAAAACAGGCTTGGCGTTATTGCTTTTAATCTCGGTAAACACCACGCCTAT 1035

Qy   1018 GATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCT 1077

Db   1036 GATGTTGGCAGTTTTCTCGATAATTACGGCATTGCTGTGCGTACCGGACATCACTGCGCA 1095

Qy   1078 CAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATAT 1137

Db   1096 ATGCCATTGATGGCCTATTACAACGTCCCTGCGATGTGTCGGGCGTCGCTGGCCATGTAT 1155

Qy   1138 AATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCG 1181

Db   1156 AACACCCATGAAGAAGTGGATCGTCTGGTGACCGGCCTGCAACG 1199

Search completed: April 22, 2003, 10:12:15
Job time : 330.905 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:19:24 ; Search time 52.6125 Seconds
(without alignments)
7029.746 Million cell updates/sec

Title: US-09-622-419-3
Perfect score: 1206
Sequence: 1 atgaacgtttttaatcccgc.....tggaattattggtggattaa 1206

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1206	100.0	3465	4	US-09-462-645C-9	Sequence 9, Appli
	2	1206	100.0	3794	4	US-09-462-645C-5	Sequence 5, Appli
	3	1205	99.9	1216	4	US-09-462-645C-1	Sequence 1, Appli
	4	192.4	16.0	18627	4	US-08-961-527-113	Sequence 113, App
	5	151.2	12.5	1287	4	US-09-134-001C-2042	Sequence 2042, Ap
	6	148	12.3	4973	4	US-09-381-862-6	Sequence 6, Appli
c	7	86.4	7.2	3447	4	US-09-221-017B-816	Sequence 816, App
c	8	86	7.1	700	4	US-09-221-017B-801	Sequence 801, App
	9	82.2	6.8	714	5	PCT-US96-05320A-1346	Sequence 1346, Ap
	10	61	5.1	1232	4	US-09-462-645C-3	Sequence 3, Appli
	11	61	5.1	3481	4	US-09-462-645C-11	Sequence 11, Appl
	12	61	5.1	3810	4	US-09-462-645C-7	Sequence 7, Appli
	13	38.2	3.2	1164	4	US-09-134-001C-788	Sequence 788, App
c	14	38	3.2	20199	4	US-08-961-527-6	Sequence 6, Appli
c	15	32.4	2.7	1068	4	US-09-403-768-5	Sequence 5, Appli
c	16	32	2.7	663	4	US-08-858-207A-97	Sequence 97, Appl
	17	31.6	2.6	1167	4	US-09-134-001C-947	Sequence 947, App
	18	31.4	2.6	222	4	US-09-367-953B-12	Sequence 12, Appl
	19	31.4	2.6	235	2	US-08-588-983-10	Sequence 10, Appl
	20	31.4	2.6	235	2	US-08-588-976-10	Sequence 10, Appl
	21	31.4	2.6	235	3	US-08-589-028-11	Sequence 11, Appl
	22	31.4	2.6	235	3	US-08-784-582-11	Sequence 11, Appl
	23	31.4	2.6	235	4	US-08-785-271-11	Sequence 11, Appl
	24	31.4	2.6	245	1	US-08-399-580B-5	Sequence 5, Appli
	25	31.4	2.6	2542	4	US-09-187-946-3	Sequence 3, Appli
	26	31.4	2.6	5470	1	US-08-441-139-12	Sequence 12, Appl
	27	31.4	2.6	5470	6	5196523-5	Patent No. 5196523
c	28	31.4	2.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
c	29	31	2.6	294	3	US-08-618-100B-1	Sequence 1, Appli
c	30	31	2.6	2921	3	US-08-618-100B-4	Sequence 4, Appli
c	31	31	2.6	3756	2	US-08-576-626A-1	Sequence 1, Appli
	32	31	2.6	4403765	4	US-09-103-840A-2	Sequence 2, Appli
	33	31	2.6	4411529	4	US-09-103-840A-1	Sequence 1, Appli
c	34	30.8	2.6	624	4	US-09-385-982-50	Sequence 50, Appl
	35	30.6	2.5	1650	4	US-09-103-754A-2	Sequence 2, Appli
	36	30.6	2.5	1766	4	US-09-326-203A-15	Sequence 15, Appl
	37	30.6	2.5	1766	4	US-09-326-203A-16	Sequence 16, Appl
	38	30.6	2.5	2381	4	US-08-628-434-3	Sequence 3, Appli
c	39	30.6	2.5	3215	4	US-09-221-017B-1096	Sequence 1096, Ap
c	40	30.4	2.5	7218	1	US-08-232-463-14	Sequence 14, Appl
	41	30.2	2.5	3228	4	US-09-605-785-701	Sequence 701, App
c	42	30.2	2.5	3478	1	US-08-530-492-1	Sequence 1, Appli
c	43	30.2	2.5	3478	4	US-08-906-517-1	Sequence 1, Appli
c	44	30.2	2.5	3484	1	US-08-530-492-105	Sequence 105, App
c	45	30.2	2.5	3484	4	US-08-906-517-105	Sequence 105, App

ALIGNMENTS

RESULT 1
US-09-462-645C-9
; Sequence 9, Application US/09462645C
; Patent No. 6436681
; GENERAL INFORMATION:
; APPLICANT: Schroeder, Hartwig
; APPLICANT: Hauer, Bernhard
; TITLE OF INVENTION: The preparation of biotin
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/462,645C
; CURRENT FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: PCT/EP98/04097
; PRIOR FILING DATE: 1998-02-07
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 9
; LENGTH: 3465
; TYPE: DNA
; ORGANISM: clone pHS2bioS1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 272...1477
US-09-462-645C-9

Query Match 100.0%; Score 1206; DB 4; Length 3465;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     272 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 331

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     332 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 391

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     392 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 451

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     452 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 511

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     512 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 571

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    572 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 631

Qy    361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    632 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 691

Qy    421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    692 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 751

Qy    481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    752 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 811

Qy    541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    812 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 871

Qy    601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    872 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 931

Qy    661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    932 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 991

Qy    721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    992 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1051

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1052 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1111

Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1112 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1171

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1172 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1231

Qy    961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1232 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1291

Qy   1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1292 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1351

Qy   1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1352 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1411

Qy   1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1412 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1471

Qy   1201 GATTAA 1206
          ||||||
Db   1472 GATTAA 1477


RESULT 2
US-09-462-645C-5
; Sequence 5, Application US/09462645C
; Patent No. 6436681
; GENERAL INFORMATION:
;  APPLICANT: Schroeder, Hartwig
;  APPLICANT:  Hauer, Bernhard
;  TITLE OF INVENTION: The preparation of biotin
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/462,645C
;  CURRENT FILING DATE:  2000-01-11
;  PRIOR APPLICATION NUMBER: PCT/EP98/04097
;  PRIOR FILING DATE: 1998-02-07
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 5
;   LENGTH: 3794
;   TYPE: DNA
;   ORGANISM: clone pHS1bioS1
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 601...1806
US-09-462-645C-5

  Query Match            100.0%;  Score 1206;  DB 4;  Length 3794;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1206;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 660

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 720

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 780

Qy    181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 840

Qy    241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 900

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 960

Qy    361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 1020

Qy    421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 1080

Qy    481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 1140

Qy    541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 1200

Qy    601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 1260

Qy    661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 1320

Qy    721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 1380

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 1440
```

```
Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 1500

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 1560

Qy    961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1620

Qy   1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1680

Qy   1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1740

Qy   1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1800

Qy   1201 GATTAA 1206
          ||||||
Db   1801 GATTAA 1806
```

RESULT 3
US-09-462-645C-1
; Sequence 1, Application US/09462645C
; Patent No. 6436681
; GENERAL INFORMATION:
; APPLICANT: Schroeder, Hartwig
; APPLICANT: Hauer, Bernhard
; TITLE OF INVENTION: The preparation of biotin
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/462,645C
; CURRENT FILING DATE: 2000-01-11
; PRIOR APPLICATION NUMBER: PCT/EP98/04097
; PRIOR FILING DATE: 1998-02-07
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 1
; LENGTH: 1216
; TYPE: DNA
; ORGANISM: E. Coli w3110
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1...11
; NAME/KEY: CDS
; LOCATION: 12...1216
US-09-462-645C-1

Query Match 99.9%; Score 1205; DB 4; Length 1216;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 71

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     72 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 131

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    132 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 191

Qy    181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    192 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 251

Qy    241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    252 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 311

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    312 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 371

Qy    361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 431

Qy    421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 491

Qy    481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    492 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 551

Qy    541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    552 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 611

Qy    601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    612 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 671

Qy    661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    672 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 731

Qy    721 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    732 GCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 791

Qy    781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    792 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 851

Qy    841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    852 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 911

Qy    901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    912 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 971

Qy    961 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    972 TTCCGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGAT 1031

Qy   1021 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1032 ATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAG 1091

Qy   1081 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1092 CCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAAT 1151

Qy   1141 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1152 ACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATTGGTG 1211

Qy   1201 GATTA 1205
          |||||
Db   1212 GATTA 1216
```

RESULT 4

```
US-08-961-527-113
; Sequence 113, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
;   APPLICANT:  Charles Kunsch
;   TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;   NUMBER OF SEQUENCES:  391
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE:  Human Genome Sciences, Inc.
;     STREET:  9410 Key West Avenue
;     CITY:  Rockville
;     STATE:  Maryland
;     COUNTRY:  USA
;     ZIP:  20850
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;     COMPUTER:  HP Vectra 486/33
;     OPERATING SYSTEM:  MSDOS version 6.2
;     SOFTWARE:  ASCII Text
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/08/961,527
;     FILING DATE:
;     CLASSIFICATION:  424
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Brookes, A. Anders
;     REGISTRATION NUMBER:  36,373
;     REFERENCE/DOCKET NUMBER:  PB340P1
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  (301) 309-8504
;     TELEFAX:  (301) 309-8512
; INFORMATION FOR SEQ ID NO:  113:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  18627 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  double
;     TOPOLOGY:  linear
US-08-961-527-113

  Query Match             16.0%;  Score 192.4;  DB 4;  Length 18627;
  Best Local Similarity   50.0%;  Pred. No. 1.6e-50;
  Matches  574;  Conservative    0;  Mismatches  556;  Indels   18;  Gaps    3;

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           |||||||| ||||  || |||||  | |  |||||   ||  ||  |  |||| |   |
Db    5253 GTCTATCTGGACAATGCTGCGACGACACAAAAACCACTAGTAGTTCTGAAAGCTATTAAC 5312

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
             |  ||    | |  |    |  || || || ||  |      | ||  |||  ||||
Db    5313 AGCTACTATGAGCAGGACAATGCCAATGTTCACCGTGGTGTCCATACCTTAGCGGAACGA 5372

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           |||  ||   ||||||||||||| || || |    |      | ||  | ||||||     |
Db    5373 GCGACAGCTTCTTATGAAGCTGCTCGTGAAACCATTCGTAAGTTTATTAATGCAGGCTCT 5432

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
             ||      |  |||  ||  | || || ||       | | |||  ||||||||   ||
Db    5433 ACAAAGGAAGTTCTCTTTACCAGAGGAACGACAACCAGCCTTAACTGGGTGGCAC---GC 5489

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
           | |||          |||       ||| ||  || |  |  |    ||    ||||||||
Db    5490 TTTGCTGAGGAAATTCTCACTGAGGGAGACCAGGTCTTGATTTCAGTAATGGAACACCAT 5549

Qy     361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
           | ||  || | || |||| |    |      | | | ||||||||  |  | ||  |  |
Db    5550 TCTAATATCATTCCATGGCAGGAAGCTTGTCGAAAGACTGGAGCAGAGCTTGTCTATGTC 5609

Qy     421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
             ||||| |   |  |      |||| | ||    |||| ||      | |||
Db    5610 TATCTTAAAG---ACGGTGCCTTGGATATGGAGGATTTGCGAGCTAAATTGACTGATAAG 5666

Qy     481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCG 540
             |   || |  |  | | |||    || || |||  |||  |    |       |   |
Db    5667 GTTAAATTTGTTTCCCTAGCTCATGCCTCCAATGTTCTTGGTGTGGTCAATCCGATCAAG 5726

Qy     541 CGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGG 600
             |  | |   || || ||   ||  |||    |  || | || ||||||||||
Db    5727 GAAATCACTCAATTAGCCCACCAAGTTGGGGCAATTATGGTAGTGGATGGTGCTCAATCT 5786

Qy     601 GCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGT 660
             ||   |||  |       ||||| ||| |  ||||| | || || |  ||| |||
Db    5787 ACACCTCATATGAAGATTGATGTCCAGGACTTGGATCTGGACTTTTTCGCCTTTTCGGGT 5846

Qy     661 CACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAG 720
           |||||  ||  ||| ||||| ||||||||  || || || || |||  | |    || |||
Db    5847 CACAAGATGGCTGGTCCGACTGGTATCGGTGTCCTTTACGGCAAAGAAAAGTATCTTGAG 5906

Qy     721 GCGATGTCGCCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTC 780
             ||||| ||          ||||||  | ||| || |    ||         |
Db    5907 CAAATGTCTCCAGTAGAATTTGGCGGCGAGATGATTGATTTTGTCTACGAGCAATTTGCT 5966

Qy     781 ACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATA 840
           |       |      ||| |||||| | || ||||||||||||||| | || || |  ||
Db    5967 AGTTGGAAGGAATTGCCTTGGAAATTTGAGGCTGGAACGCCAAATATGGCAGGAGCTATT 6026

Qy     841 GGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGG 900
           |||  |     | ||  | ||  |  |||| | |     | |||  |    |  |||
Db    6027 GGACTTGCGACTGCAGTTGATTATCTGGAAAAGATTGGTATGGATGCCGTTGAAGCTCAT 6086

Qy     901 AGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCA 960
             |     ||     ||   |   |        |||  |  |  |   || ||
Db    6087 GAACAGGAATTGATTGCGTACGTCTATCCAAAACTGCAGGCAATTGAGGGATTGACCATT 6146

Qy     961 TTCCGCTGCCAGGA-----------TTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTT 1008
           | | | |  |||||           ||  |  |  | ||||||  |  | | ||   |
Db    6147 TACGGTTCTCAGGATTTGGCTCAACGTTCGGGTGTTATTGCCTTTAACCTAGGTGATCTC 6206

Qy    1009 CATCATAGCGATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAG 1068
           ||||  |  |||| | | ||||   ||||   |  | ||  | ||  |  || || || ||
Db    6207 CATCCTCACGATCTTGCGACGGCTCTGGATTATGAAGGAGTGGCTGTTCGTGCTGGTCAC 6266

Qy    1069 CATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTT 1128
           ||||| || || ||  | ||     | || |  ||  | |  |||   || ||   ||||
Db    6267 CATTGTGCGCAACCCTTGCTTCAGTATTTGGAAGTCCCAGCAACAGCTCGTGCAAGTTTT 6326

Qy    1129 GCGCCATATAATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTG 1188
             || ||||| |||   |||   ||    || ||  ||||| |  |      |  |
Db    6327 TATATCTACAATACCAAGGCAGATTGCGACAAACTAGTCGATGCCCTACAAAAGACAAAG 6386

Qy    1189 GAATTATT 1196
           || || ||
Db    6387 GAGTTTTT 6394


RESULT 5
US-09-134-001C-2042
; Sequence 2042, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOC
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 2042
;   LENGTH: 1287
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
```

US-09-134-001C-2042

Query Match 12.5%; Score 151.2; DB 4; Length 1287;
Best Local Similarity 46.4%; Pred. No. 5e-38;
Matches 535; Conservative 0; Mismatches 608; Indels 9; Gaps 1;

```
Qy     64 TATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAACAG 123
Db    130 TACCTTGATTCAACTGCGACTAGCCAAACACCTGTACAAGTTTTAAATGTACTAGATGAT 189

Qy    124 TTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGCCTG 183
Db    190 TACTACAAACGCTATAATTCTAATGTCCATCGTGGTGTACACACATTAGGGTCTTTAGCT 249

Qy    184 ACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGAT 243
Db    250 ACAGATGGCTATGAAAATGCAAGAGAAACAGTAAGACGTTTTATCAATGCGAAGTACTTT 309

Qy    244 AAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTAT 303
Db    310 GAAGAAATTATTTTCACTCGAGGAACAACTGCATCAATAAATATAGTAGCACATAGTTAT 369

Qy    304 GCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCC 363
Db    370 GGTGACGCTAATATTAGTGAGGGTGACGAAATTGTTGTAACAGAAATGGAACATCATGCT 429

Qy    364 AACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCG 423
Db    430 AATATTGTTCCATGGCAACAATTAGCTAAACGTAAAAATGCAACACTAAAATTTATTCCA 489

Qy    424 CTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGT 483
Db    490 ATGACAAAAGATGGTGAATTGCAATTAGATGATATAAAAGCAACCATCAATGATAAAACT 549

Qy    484 CGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGA 543
Db    550 AAAATTGTTGCTATCGCACACGTTTCCAATGTTTTAGGAACAATTAATGATGTTAAAACT 609

Qy    544 GCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCA 603
Db    610 ATCGCTAAAATTGCACATGAACATGGTGCAGTTATAAGTGTGGATGGTGCGCAATCAGCT 669

Qy    604 GTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCAC 663
Db    670 CCACACATGGCTTTAGATATGCAAGATATCGATGCTGATTTTTATAGTTTTAGTGGACAT 729

Qy    664 AAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCG 723
Db    730 AAAATGTTAGGTCCTACTGGCATTGGAGTTTTATATGGAAAAAGAGAATTATTACAAAAT 789

Qy    724 ATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACG 783
Db    790 ATGGAACCCGTTGAATTTGGTGGAGATATGATAGACTTCGTAAGTAAATATGACGCCACT 849

Qy    784 ACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGA 843
Db    850 TGGGCTGACCTACCAACTAAATTTGAGGCTGGCACACCACTTATTGCTCAAGCAATAGGT 909

Qy    844 TTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGGAGC 903
Db    910 TTAGCAGAAGCTATACGTTATATTGAAAATTTAGGTTTTAATGCGATTCATCAGCATGAA 969

Qy    904 CGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCATTC 963
Db    970 AAAGAATTGACAGAATATGCATATGAACAAATGTTAACTATAGATGGTCTTGAAATTTAC 1029

Qy    964 CGCTGCCAGGATTCCAGCC---------TGCTGGCCTTTGATTTTGCTGGCGTTCATCAT 1014
Db   1030 GGTCCACCAAAAGATAGACGTGCAGGGGTCATTACTTTTAATTTAGCAGATATACATCCA 1089

Qy   1015 AGCGATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGC 1074
Db   1090 CATGATGTAGCTACAGCTGTTGATACAGAAGGTGTTGCAGTTCGTGCAGGCCATCACTGT 1149

Qy   1075 GCTCAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCA 1134
Db   1150 GCGCAACCTTTGATGAAATGGTTAGGTGTTTCTTCAACAGCACGTGCAAGTTTTTACGTT 1209

Qy   1135 TATAATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTA 1194
Db   1210 TATAATACGAAAGAAGATGTAGATCAATTAGTTCAAGCATTGAAACAAACGAAGGAGTTT 1269

Qy   1195 TTGGTGGATTAA 1206
Db   1270 TTCTCTTATGAA 1281
```

RESULT 6
US-09-381-862-6
; Sequence 6, Application US/09381862
; Patent No. 6245906
; GENERAL INFORMATION:
; APPLICANT: Ueyama, Hiroshi
; APPLICANT: Abe, Kanako
; APPLICANT: Keshi, Hiroyuki
; APPLICANT: Matsuhisa, Akio
; TITLE OF INVENTION: PROBES FOR THE DIAGNOSIS OF INFECTIONS
; TITLE OF INVENTION: CAUSED BY STREPTOCOCCUS PYOGENES
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/381,862
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 1997-71077
; FILING DATE: 25-MAR-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP98/01288
; FILING DATE: 23-MAR-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Cawley, Jr., Thomas A.
; REGISTRATION NUMBER: 40,944
; REFERENCE/DOCKET NUMBER: 19036/36274
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4973 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pyogenes
; STRAIN: Clinical Isolate SP-55-3
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1934
; OTHER INFORMATION: /note= "N = adenine or cytosine or
; OTHER INFORMATION: guanine or thymine"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 3740

```
;      OTHER INFORMATION:  /note= "N = adenine or cytosine or
;      OTHER INFORMATION:  guanine or thymine"
US-09-381-862-6

  Query Match             12.3%;  Score 148;  DB 4;  Length 4973;
  Best Local Similarity   48.7%;  Pred. No. 1e-36;
  Matches  559;  Conservative    0;  Mismatches  571;  Indels   19;  Gaps    5;

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           || ||| | || |  ||||| ||| | |  |||||      ||  |||| ||    |||
Db    1147 GTTTATTTGGATAATGCCGCCACCACACAAAAACCGGCGCTTGTTCTTGAGGCTTTGCAA 1206

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
             | |||              |  || ||||| ||  |      | ||   ||  ||||
Db    1207 TCCTATTATCAAGAAGATAATGCTAATGTCCACCGAGGAGTTCATACCTTGGCTGAACGT 1266

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
             ||       ||||| ||    ||  || | || ||  | ||  |  ||||
Db    1267 GCAACGCACAAATATGAGGCCAGTCGCCAGCAGGTTGCTGACTTTATTCATGCTAAATCA 1326

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
           |||      |  |||  ||  | || || ||         | ||   ||| || |   |
Db    1327 AGTAAGGAAGTGCTCTTTACCAGAGGAACAACAACCAGTTTGAATTGGGTTGCTC---GG 1383

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
           | |||    | |    ||   || |  |||||| |  |  |      |    || || ||
Db    1384 TTTGCAGAGCAGGTCTTGACGCCAGAAGATGAGGTGTTGATTTCGATTATGGAGCATCAT 1443

Qy     361 GCCAACCTC-GTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATT 419
           |||||  ||  |||  ||||     | |   ||| |||| ||||| |   | ||  |  |
Db    1444 GCCAATATCATTCCTTTGGCAACAAGCCTGTCAAAAAACAGGAGCAAGGTTAGTCTATGT 1503

Qy     420 GCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCG 479
             | || |   |  | |  |  ||  | ||    ||| || |     | ||  |
Db    1504 TTATTTAAAAG---ATGGCCAACTTGACATGGACGATTTGGCAAACAAATTGACGACAAA 1560

Qy     480 TAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGC 539
             |  ||  || |     | |  ||  | || || |||  |||  |   |       |
Db    1561 AACACGTTTTGTTAGCCTAGTACATGTCTCCAATGTTCTTGGTTGCATCAATCCCATTAA 1620

Qy     540 GCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGG 599
             |      |  | || || |    ||        |  | ||||| ||||| |||
Db    1621 AGAAATTGCCAAGCTGGCACATGCTAAAGGAGCCTACCTTGTTGTTGACGGTGCCCAGTC 1680

Qy     600 GGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGG 659
           ||     |||||  |    |||||||  |  |||||  |||||||| ||||||||||
Db    1681 GGTTCCACATTTGGCTATTGATGTTCAAGACTTGGATTGTGATTTCTTTGCTTTTTCAGC 1740

Qy     660 TCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGA 719
           |||  ||  |||  || || |||||| | || || || || || |||  ||| ||  || |
Db    1741 TCATAAGATGTTGGGGCCAACAGGTTTGGGTGTTCTTTACGGCAAAGAAGAGCTTTTGAA 1800

Qy     720 GGCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTT 779
             ||  |||           |||||  ||||| || |    ||        | |
Db    1801 TCAAGTGGAGCCTCTTGAATTTGGCGGAGAAATGATTGATTTTGTTTACGAACAAGAGGC 1860

Qy     780 CACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCAT 839
           |||     ||     ||| |||||  | ||||| ||||| ||  |  | |||||||  ||
Db    1861 CACTTGGAAAGAATTGCCCTGGAAGTTTGAAGCAGGAACACCTCACATAGCTGGTGCTAT 1920

Qy     840 AGGATTAAGCGCGGC---------GCTGGAATGGCTGGCAGATTACGATATCAACCAGGC 890
           ||  |||||||  |           ||  |  | || |         |||||  |   |
Db    1921 TGGGCTAAGCGCANCCATTTCTTACCTTCAGAGACTAGGCATGGCTGATATACATGCGCA 1980

Qy     891 CGAAAGCTGGAGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGG 950
           |||              ||  |    ||  |      ||| ||  | | |   | | ||
Db    1981 TGAAGCAGAACTAATAGCCTATGTCTTGCCGAAATTAGAAGCTATTGAGGGGCTTACCAT 2040

Qy     951 CTTTCGTTCATTC---CGCTGCCAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGT 1007
           | | |  ||  |   |   |        ||  | ||| |  | ||| || | | ||   |
Db    2041 ATATGGACCAAGCCAGCCTAGTGCAAGATCTGGTCTGATTTCTTTTAATCTGGATGATTT 2100

Qy    1008 TCATCATAGCGATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCA 1067
             |||| |   ||  |||  ||     |||   |  | ||| |||||  |  | || |||||
Db    2101 GCATCCTCATGACTTGGCAACAGCCTTGGACTATGAAGGTGTTGCAGTAAGAGCAGGGCA 2160

Qy    1068 GCATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTT 1127
             || ||||| || || || ||     | ||||| ||| | |  ||  |  | ||   |||
Db    2161 CCACTGCGCCCAACCTCTTCTTAGTTATTTAGGTGTACCAGCAACTGTTAGAGCAAGTTT 2220

Qy    1128 TGCGCCATATAATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCT 1187
           |      ||||| || |||   |||   ||    || || || ||  ||      ||
Db    2221 TTATATCTATAACACCAAGGCAGATTGTGACCGTCTAGTCGAAGCAATTCTAAAAGCAAA 2280

Qy    1188 GGAATTATT 1196
           ||| || ||
Db    2281 GGAGTTTTT 2289


RESULT 7
US-09-221-017B-816/c
; Sequence 816, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  816:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  3447 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
```

```
;      LOCATION:    1...3447
US-09-221-017B-816

  Query Match              7.2%;  Score 86.4;  DB 4;  Length 3447;
  Best Local Similarity   49.3%;  Pred. No. 2.9e-17;
  Matches  225;  Conservative    0;  Mismatches  231;  Indels    0;  Gaps    0;

Qy      60 CGTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCA 119
           | ||||||| |||| ||||| || | | |||||   ||| ||  | ||    | | |
Db     458 CATCTATCTGGACAATGCCGCCACGACACAGAAACCCATAGCGGTATTAGACCAAATCGA 399

Qy     120 ACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACG 179
           ||     ||| |      |   |  ||  |||| ||| ||                ||
Db     398 ACGAGCCTACCGCACATTCAATGCCAATATCCACCGCGGCGTACACCATCTGAGTCAGGT 339

Qy     180 CCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGA 239
               || | ||   | |||  ||| ||  ||    | || |  |   |    ||    ||
Db     338 GGCTACGGAGCAACACGAAAATGCTCGCAAGCGGATAGCCCGCTACATCGGAGCGGAAGA 279

Qy     240 TGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATG 299
              |||    |  |||  ||||||||||||||| ||  | |||||  |||||||    |
Db     278 CCCCAAAGAGGTGCTCTTTACGCGCGGCACCACAGAGGCAATCAATCTGGTGGCCTCTTC 219

Qy     300 CTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCA 359
           | | |  |     | |   | ||||||||||||||  ||||       ||||||||
Db     218 GTTCGGCCAAGCTTTCGTCCGCTCCGGCGATGAGATTATCATCAGTACCCTCGAACACCA 159

Qy     360 CGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATT 419
             | ||  |||||||  ||||   |  | |||  |||  || || ||      |       |
Db     158 TTCGAATATCGTCCCATGGCAACTTTTGGCCGAACGTACCGGTGCTCGGCTCAGAGTGAT 99

Qy     420 GCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCG 479
           | | | |   |  | |    | ||||||  ||    || |     ||  | |     ||
Db      98 CCGGTTGACCGAATCGGGCGAACTGGATGTGGAGCATTTCCGGTCTCTACTGAACGAACG 39

Qy     480 TAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGT 515
           ||     ||  |  || | |  ||  |    || ||
Db      38 TACCAAAATCGTTTCGATAGCCCACGTCAGCAATGT 3


RESULT 8
US-09-221-017B-801/c
; Sequence 801, Application US/09221017B
; Patent No. 6444799
;  GENERAL INFORMATION:
;    APPLICANT:  Ross, Bruce C.
;    TITLE OF INVENTION:  P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
;    NUMBER OF SEQUENCES:  1120
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  MORRISON & FOERSTER
;      STREET:  755 PAGE MILL ROAD
;      CITY:  Palo Alto
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94304-1018
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  Windows
;      SOFTWARE:  FastSEQ for Windows Version 2.0b
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/221,017B
;      FILING DATE:  23-DEC-1998
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1182
;      FILING DATE:  31-DEC-1997
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP1546
;      FILING DATE:  30-JAN-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PP2911
;      FILING DATE:  09-APR-1998
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/AU98/01023
;      FILING DATE:  10-DEC-1998
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Monroy, Gladys H
;      REGISTRATION NUMBER:  32,430
;      REFERENCE/DOCKET NUMBER:  27340-20021.00
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-813-5600
;      TELEFAX:  650-494-0792
;      TELEX:  706141
;  INFORMATION FOR SEQ ID NO:  801:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  700 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  circular
;    MOLECULE TYPE:  DNA (genomic)
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  UNKNOWN
;    ORIGINAL SOURCE:
;      ORGANISM:  PORYPHYROMONAS GINGIVALIS
;    FEATURE:
;      NAME/KEY:  misc_feature
;      LOCATION:  1...700
US-09-221-017B-801

  Query Match              7.1%;  Score 86;  DB 4;  Length 700;
  Best Local Similarity   49.6%;  Pred. No. 1.8e-17;
  Matches  279;  Conservative    0;  Mismatches  276;  Indels    8;  Gaps    2;

Qy     627 GCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTAT 686
           | || | |  |  |||||||| || || || || || ||| | || ||| | ||||| ||
Db     686 GGAAGTTGGAACGGATTTCTACGCCTTCTCCGGCCATAAAGTCTACGGCACCACAGGAAT 627

Qy     687 CGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGG 746
            |||||  ||   |  ||    | | || | || |       | ||||  |  || |||||
Db     626 AGGCGTATTGATGGCAAAGC--GNAGTGGTTGAAGAAGATCCCCCCTACCACGGTGGCGG 569

Qy     747 CAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACT 806
             | ||| |   | | ||||  || ||       |||      |      ||| |  || |
Db     568 AGAGATGATCGCCAAGGTGACATTCGAAAAGACCACTTACAACGAATTGCCCTATAAGTT 509

Qy     807 GGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCT 866
             ||||| || |||||  |    |  ||  ||  |  ||   | | || || || |   |
Db     508 CGAAGCCGGCACGCCGGACTATGTCGGCTCCACGGCTTTGGCCACAGCTCTCGACTACAT 449

Qy     867 GGCAGATTACGATATCAACCAGGCCGAAAGCTGGAGCCGTAGCTTAGCAACGCTGGCGGA 926
            | ||  |  | | |  |      |                |  |      ||||
Db     448 CGAAGGATTGGGTTTGGAAGCCATAGCCGCACATGAAGACAAACTCCTGCGCATGGCCAC 389

Qy     927 AGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCATTCCG------CTGCCAGGATTCCAG 980
            || |   ||          |||      |||    || |      |   || |   | |
Db     388 CGAAGGCGTGAGAGCTTTCCCCAATGCCCGTATCATCGGAGAAGCCGAACACAAGAGCGG 329

Qy     981 CCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGATATGGTGACGCTGCTGGCGGA 1040
           ||||  | ||   | | |  | | |||||||     ||  | |  | |||||| |
Db     328 TGTGCTCTCTTTCCTTATAGGGGACATTCATCCCTATGACCTCGGCATGCTGCTCGATCG 269

Qy    1041 GTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAGCCGCTACTGGCAGAATTAGG 1100
           |   ||||| ||| | ||  |||| || || || ||  | |||||||| | |    | ||
Db     268 GCTGGGTATAGCCATCCGTACCGGACATCACTGTGCCGAACCGCTACTCGAAAGCATGGG 209

Qy    1101 CGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAATACAAAGAGTGATGTGGATGC 1160
             | |  | |||    |||||||| ||||| |  || |||||  ||   || |||||
Db     208 ACTCAATGCCACGGCTCGCGCCTCCTTTGCTCTCTACAATACCGAGGAGGAAGTGGAGCA 149

Qy    1161 GCTGGTGAATGCCGTTGACCGCG 1183
             |  |  | |||  |  | || |
Db     148 ATTCCTCCAAGCCTTGAAACGTG 126


RESULT 9
```

```
PCT-US96-05320A-1346
; Sequence 1346, Application PC/TUS9605320A
;  GENERAL INFORMATION:
;     APPLICANT:  Human Genome Sciences
;     APPLICANT:  9410 Key West Avenue
;     APPLICANT:  Rockville, MD 20850
;     APPLICANT:  United States of America
;     APPLICANT:  Johns Hopkins University
;     APPLICANT:  720 Rutland Avenue
;     APPLICANT:  Baltimore, MD 21205
;     APPLICANT:  United States of America
;     APPLICANT:  Mark D. Adams
;     APPLICANT:  Owen White
;     APPLICANT:  Hamilton O. Smith
;     APPLICANT:  J. Craig Venter
;     TITLE OF INVENTION:  Nucleotide Sequence of the Haemophilus Influenzae Rd Genome,
;     NUMBER OF SEQUENCES:  48
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Sterne, Kessler, Goldstein & Fox
;       STREET:  1100 New York Avenue, Suite 600
;       CITY:  Washington
;       STATE:  D.C.
;       COUNTRY:  USA
;       ZIP:  20003-3934
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;       COMPUTER:  HP Vectra 486/33
;       OPERATING SYSTEM:  MSDOS version 6.2
;       SOFTWARE:  ASCII Text
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  PCT/US96/05320A
;       FILING DATE:  April22, 1996
;       CLASSIFICATION:
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/476,102
;       FILING DATE:  June 7, 1995
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  08/487,429
;       FILING DATE:  June 7, 1995
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Eric K. Steffe
;       REGISTRATION NUMBER:  36,688
;       REFERENCE/DOCKET NUMBER:  1488.014PC01
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (202) 371-2600
;       TELEFAX:  (202) 371-2540
;   INFORMATION FOR SEQ ID NO:  1346:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  714 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  double
;       TOPOLOGY:  linear
PCT-US96-05320A-1346

  Query Match             6.8%;  Score 82.2;  DB 5;  Length 714;
  Best Local Similarity   48.1%;  Pred. No. 2.8e-16;
  Matches  265;  Conservative    0;  Mismatches  283;  Indels    3;  Gaps    1;

Qy     347 TGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCA 406
           |||   | ||    || ||  | ||| | ||||         ||  || |    || ||'|
Db     116 TGGTTCATCAAGTTGCTAATTTTGTCACTTGGCATGAAACGGCGAAAAAGTGCGGTGCAA 175

Qy     407 AAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAAC 466
           || |       || ||| ||   |   | | ||    ||||   || || |
Db     176 AAATTCGAGTTTTACCGATTTTAGATAATTGGCTAATTGATGAAAATGTGCTTATTTCTA 235

Qy     467 TGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTT 526
             || ||      |       |  | ||| ||  |    | || || || ||||||
Db     236 CCCTTTCTGAAAAAACAAAACTCGTTGCGCTGAATTTTGTTTCAAATGTCACTGGCACAG 295

Qy     527 GCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCGG---GATGGTGGTGATGG 583
           | |   |    ||  ||||   || |     |    |     | ||||  | |
Db     296 AACAGCCGATTAAACGCCTGATTCAACTTATTAGAAAACATAGTCATGCGTTGGTTTTAG 355

Qy     584 TTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATT 643
           | ||||  || || | |      ||| |        ||| | ||  |  | |||  ||||
Db     356 TGGATGCAGCACAAGCGATTAGTCATATTAAAATTGATTTACAAGATTTAGATGCCGATT 415

Qy     644 TCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTA 703
           ||   || ||||| |  || ||| | |||||||| |  || | ||||  |    || |
Db     416 TCCTCGCATTTTCTGCCCATAAAATTTATGGCCCAAATGGACTTGGCGTTTTAACAGGAA 475

Qy     704 AATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAG 763
           |||         ||        |   ||      |   ||||| |||||||| |    |
Db     476 AATTGACCGCACTTTCTCAACTTCAACCGCTCTTTTTTGGCGGAAAAATGGTTGAACGTG 535

Qy     764 TGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAA 823
           ||      | |  |  | ||  |     ||| |     | |||||||| || || |
Db     536 TGTCAAATAATCGTATTACTTTTGCCGAATTGCCTTATCGTTTAGAAGCTGGCACACCCA 595

Qy     824 ATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCA 883
           || | ||||| || ||||| || |  || | ||||| |||||   | | |  ||| | |
Db     596 ATATTGCTGGGGTTATAGGCTTTAATGCTGTGCTGGATTGGCTACAAAAATGGGATTTTA 655

Qy     884 ACCAGGCCGAA 894
           |   || |||
Db     656 CCGCAGCAGAA 666


RESULT 10
US-09-462-645C-3
; Sequence 3, Application US/09462645C
; Patent No. 6436681
; GENERAL INFORMATION:
;  APPLICANT: Schroeder, Hartwig
;  APPLICANT:  Hauer, Bernhard
;  TITLE OF INVENTION: The preparation of biotin
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/462,645C
;  CURRENT FILING DATE:  2000-01-11
;  PRIOR APPLICATION NUMBER: PCT/EP98/04097
;  PRIOR FILING DATE: 1998-02-07
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 3
;   LENGTH: 1232
;   TYPE: DNA
;   ORGANISM: E. Coli w3110
;   FEATURE:
;   NAME/KEY: 5'UTR
;   LOCATION: 1...18
;   NAME/KEY: CDS
;   LOCATION: 19...1232
US-09-462-645C-3

  Query Match             5.1%;  Score 61;  DB 4;  Length 1232;
  Best Local Similarity   46.1%;  Pred. No. 2e-09;
  Matches  242;  Conservative    0;  Mismatches  280;  Indels    3;  Gaps    1;

Qy     250 ATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCGT 309
           |||||||  ||    ||  | || ||||   |||| ||| |    || |  ||
Db     223 ATCGTCTTTACCTCTGGTGCAACCGAATCTGACAACCTGGCGATCAAAGGTGCAGCCAAC 282

Qy     310 CCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTC 369
             || | ||   ||| |  | || ||   ||||    | |||||| | ||    ||
Db     283 TTTTATCAGAAAAAAGGCAAGCACATCATCACCAGCAAAACCGAACACAAAGCGGTACTG 342

Qy     370 GTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAAT 429
           |   |||| |    |: | |  |  ||   ||     ||||     |  || | |   |
Db     343 GATACCTGCCGTCAGCTGGAGCGCGAA---GGTTTTGAAGTCACCTACCTGGCACCGCAG 399

Qy     430 GCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATT 489
           |  | |   ||  | |      |   || |  |||  |  |  |      |
Db     400 CGTAACGGCATTATCGACCTGAAAGAACTTGAAGCAGCGATGCGTGACGACACCATCCTC 459

Qy     490 CTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATT 549
```

```
Db    460 GTGTCCATCATGCACGTAAATAACGAAATCGGCGTGGTGCAGGATATCGCGGCTATCGGC 519

Qy    550 ACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCAT 609

Db    520 GAAATGTGCCGTGCTCGTGGCATTATCTATCACGTTGATGCAACCCAGAGCGTGGGTAAA 579

Qy    610 TTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTG 669

Db    580 CTGCCTATCGACCTGAGCCAGTTGAAAGTTGACCTGATGTCTTTCTCCGGTCACAAAATC 639

Qy    670 TATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCG 729

Db    640 TATGGCCCGAAAGGTATCGGTGCGCTGTATGTACGTCGTAAACCGCGCGTACGCATCGAA 699

Qy    730 CCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGAC 774

Db    700 GCGCAAATGCACGGCGGCGGTCACGAGCGCGGTATGCGTTCCGGC 744


RESULT 11
US-09-462-645C-11
; Sequence 11, Application US/09462645C
; Patent No. 6436681
; GENERAL INFORMATION:
;  APPLICANT: Schroeder, Hartwig
;  APPLICANT:  Hauer, Bernhard
;  TITLE OF INVENTION: The preparation of biotin
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/462,645C
;  CURRENT FILING DATE:  2000-01-11
;  PRIOR APPLICATION NUMBER: PCT/EP98/04097
;  PRIOR FILING DATE: 1998-02-07
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 11
;   LENGTH: 3481
;   TYPE: DNA
;   ORGANISM: clone pHS2bioS2
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 279...1493
US-09-462-645C-11

  Query Match             5.1%;  Score 61;  DB 4;  Length 3481;
  Best Local Similarity   46.1%;  Pred. No. 3.3e-09;
  Matches  242;  Conservative    0;  Mismatches  280;  Indels    3;  Gaps    1;

Qy    250 ATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCGT 309

Db    483 ATCGTCTTTACCTCTGGTGCAACCGAATCTGACAACCTGGCGATCAAAGGTGCAGCCAAC 542

Qy    310 CCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTC 369

Db    543 TTTTATCAGAAAAAAGGCAAGCACATCATCACCAGCAAAACCGAACACAAAGCGGTACTG 602

Qy    370 GTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAAT 429

Db    603 GATACCTGCCGTCAGCTGGAGCGCGAA---GGTTTTGAAGTCACCTACCTGGCACCGCAG 659

Qy    430 GCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATT 489

Db    660 CGTAACGGCATTATCGACCTGAAAGAACTTGAAGCAGCGATGCGTGACGACACCATCCTC 719

Qy    490 CTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATT 549

Db    720 GTGTCCATCATGCACGTAAATAACGAAATCGGCGTGGTGCAGGATATCGCGGCTATCGGC 779

Qy    550 ACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCAT 609

Db    780 GAAATGTGCCGTGCTCGTGGCATTATCTATCACGTTGATGCAACCCAGAGCGTGGGTAAA 839

Qy    610 TTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTG 669

Db    840 CTGCCTATCGACCTGAGCCAGTTGAAAGTTGACCTGATGTCTTTCTCCGGTCACAAAATC 899

Qy    670 TATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCG 729

Db    900 TATGGCCCGAAAGGTATCGGTGCGCTGTATGTACGTCGTAAACCGCGCGTACGCATCGAA 959

Qy    730 CCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGAC 774

Db    960 GCGCAAATGCACGGCGGCGGTCACGAGCGCGGTATGCGTTCCGGC 1004


RESULT 12
US-09-462-645C-7
; Sequence 7, Application US/09462645C
; Patent No. 6436681
; GENERAL INFORMATION:
;  APPLICANT: Schroeder, Hartwig
;  APPLICANT:  Hauer, Bernhard
;  TITLE OF INVENTION: The preparation of biotin
;  FILE REFERENCE:
;  CURRENT APPLICATION NUMBER: US/09/462,645C
;  CURRENT FILING DATE:  2000-01-11
;  PRIOR APPLICATION NUMBER: PCT/EP98/04097
;  PRIOR FILING DATE: 1998-02-07
;  NUMBER OF SEQ ID NOS: 38
;  SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 7
;   LENGTH: 3810
;   TYPE: DNA
;   ORGANISM: clone pHS1bioS2
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 608...1822
US-09-462-645C-7

  Query Match             5.1%;  Score 61;  DB 4;  Length 3810;
  Best Local Similarity   46.1%;  Pred. No. 3.5e-09;
  Matches  242;  Conservative    0;  Mismatches  280;  Indels    3;  Gaps    1;

Qy    250 ATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCGT 309

Db    812 ATCGTCTTTACCTCTGGTGCAACCGAATCTGACAACCTGGCGATCAAAGGTGCAGCCAAC 871

Qy    310 CCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTC 369

Db    872 TTTTATCAGAAAAAAGGCAAGCACATCATCACCAGCAAAACCGAACACAAAGCGGTACTG 931

Qy    370 GTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAAT 429

Db    932 GATACCTGCCGTCAGCTGGAGCGCGAA---GGTTTTGAAGTCACCTACCTGGCACCGCAG 988

Qy    430 GCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATT 489

Db    989 CGTAACGGCATTATCGACCTGAAAGAACTTGAAGCAGCGATGCGTGACGACACCATCCTC 1048

Qy    490 CTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATT 549

Db   1049 GTGTCCATCATGCACGTAAATAACGAAATCGGCGTGGTGCAGGATATCGCGGCTATCGGC 1108

Qy    550 ACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCAT 609

Db   1109 GAAATGTGCCGTGCTCGTGGCATTATCTATCACGTTGATGCAACCCAGAGCGTGGGTAAA 1168

Qy    610 TTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTG 669

Db   1169 CTGCCTATCGACCTGAGCCAGTTGAAAGTTGACCTGATGTCTTTCTCCGGTCACAAAATC 1228

Qy    670 TATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCG 729

Db   1229 TATGGCCCGAAAGGTATCGGTGCGCTGTATGTACGTCGTAAACCGCGCGTACGCATCGAA 1288

Qy    730 CCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGAC 774
```

```
Db   1289 GCGCAAATGCACGGCGGCGGTCACGAGCGCGGTATGCGTTCCGGC 1333


RESULT 13
US-09-134-001C-788
; Sequence 788, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;  APPLICANT: Lynn Doucette-Stamm et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
;  TITLE OF INVENTION:  EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC-007
;  CURRENT APPLICATION NUMBER: US/09/134,001C
;  CURRENT FILING DATE:  1998-08-13
;  PRIOR APPLICATION NUMBER: US 60/064,964
;  PRIOR FILING DATE: 1997-11-08
;  PRIOR APPLICATION NUMBER: US 60/055,779
;  PRIOR FILING DATE: 1997-08-14
;  NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 788
;   LENGTH: 1164
;   TYPE: DNA
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-788

  Query Match             3.2%;  Score 38.2;  DB 4;  Length 1164;
  Best Local Similarity   56.9%;  Pred. No. 0.033;
  Matches   70;  Conservative    0;  Mismatches   53;  Indels    0;  Gaps    0;

Qy     583 GTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGAT 642
           ||||||| || ||| |     | |||||       ||| ||||  |  |  | ||||||
Db     541 GTTGATGCTGTTCAAGCAATTGGACATTTAGATTTAGATTTTCATAATTTTAAAATTGAT 600

Qy     643 TTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGT 702
               | ||||||  |||||| |   ||| || | |||| | ||  | || |  |
Db     601 ACAATGAGTATTTCAGCACACAAATTTGGTGGTCCTAAAGGTGTTGGTCTACTATTAGTA 660

Qy     703 AAA 705
           |||
Db     661 AAA 663


RESULT 14
US-08-961-527-6/c
; Sequence 6, Application US/08961527
; Patent No. 6420135
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Streptococcus pneumoniae Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  391
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,527
;      FILING DATE:
;      CLASSIFICATION:  424
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  6:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  20199 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-961-527-6

  Query Match             3.2%;  Score 38;  DB 4;  Length 20199;
  Best Local Similarity   57.6%;  Pred. No. 0.16;
  Matches   68;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy     583 GTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGAT 642
           |||||||  | |||||     |   |  ||||     ||    |  || |||  ||||||
Db   13792 GTTGATGCAGTTCAGGCTATTGGTAAAATCCCAATTCATTCAGAAGAATTGGGCATTGAT 13733

Qy     643 TTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATG 700
           ||      ||  ||| |  |||||| |  '||| || |  ||  |||||  | || || |
Db   13732 TTTCTCACTGCTTCTGCCCACAAATTCCATGGTCCTAAGGGAATCGGTTTTCTCTACG 13675


RESULT 15
US-09-403-768-5/c
; Sequence 5, Application US/09403768
; Patent No. 6444804
; GENERAL INFORMATION:
;  APPLICANT:  Lam, Joseph S.
;  APPLICANT:  De Kievit, Teresa R.
;  APPLICANT:  Burrows, Lori L.
;  APPLICANT:  Walsh, Andrew
;  APPLICANT:  Matewish, Mauricia
;  TITLE OF INVENTION:  No. 6444804el Proteins Involved in the Synthesis and Assembl
;  TITLE OF INVENTION:  of Core Lipopolysaccharide of Pseudomonas Aeruginosa
;  FILE REFERENCE:  6580-177
;  CURRENT APPLICATION NUMBER: US/09/403,768
;  CURRENT FILING DATE:  1999-11-02
;  PRIOR APPLICATION NUMBER:  U.S. 60/045,418
;  PRIOR FILING DATE:  1997-05-02
;  PRIOR APPLICATION NUMBER:  U.S. 60/046,149
;  PRIOR FILING DATE:  1997-05-09
;  NUMBER OF SEQ ID NOS:  8
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 1068
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa rfaC (WaaC)
;   FEATURE:
;   OTHER INFORMATION: Description of Organism: Pseudomonas aeruginosa
;   OTHER INFORMATION: serotype O5 strain PAO1
US-09-403-768-5

  Query Match             2.7%;  Score 32.4;  DB 4;  Length 1068;
  Best Local Similarity   50.6%;  Pred. No. 2.2;
  Matches   78;  Conservative    0;  Mismatches   76;  Indels    0;  Gaps    0;

Qy     323 CGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGA 382
           | |||||   |  | ||| |  ||| || |  || |  | ||  ||    |     |||
Db     338 CCGGCGACCGGCGTCTTGCCCACGTAGCGGGTCAGCCAGGCACTCTTCAGCAGCCCCTGG 279

Qy     383 TGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGC 442
             ||||  || ||      || | |  |  |  |  || |||| | |||| || | |
Db     278 GCGTCGATCACCAGGTCATAGTCGACTTCCTTCAGGCGCTGCTTGAAGCGCCGCCATTCG 219

Qy     443 CGGATGTCGATTTGTTGCCAGAACTGATTACTCC 476
           | | ||   |     |||||||  |  || | ||
Db     218 CCGTTGCGCAGGGTCTGCCAGAGGTTCTTGCGCC 185


Search completed: April 22, 2003, 14:27:34
```

Job time : 134.612 secs

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 12:23:08 ; Search time 154.262 Seconds
(without alignments)
8506.861 Million cell updates/sec

Title: US-09-622-419-3
Perfect score: 1206
Sequence: 1 atgaacgtttttaatcccgc.....tggaattattggtggattaa 1206

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1419640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	234.6	19.5	1217	10	US-09-974-300-4739	Sequence 4739, Ap
	2	234.6	19.5	1224	10	US-09-974-300-4738	Sequence 4738, Ap
	3	198.4	16.5	10397	10	US-09-070-927A-290	Sequence 290, App
c	4	175.2	14.5	4280	7	US-08-781-986A-81	Sequence 81, Appl
	5	145.2	12.0	727	10	US-09-974-300-304	Sequence 304, App
c	6	138	11.4	265	10	US-09-815-242-3775	Sequence 3775, Ap
	7	119.4	9.9	1263	9	US-09-738-626-1727	Sequence 1727, Ap
	8	72.6	6.0	419	10	US-09-878-574-3815	Sequence 3815, Ap
	9	58.8	4.9	246	10	US-09-974-300-449	Sequence 449, App
c	10	55.6	4.6	162	9	US-09-738-626-1726	Sequence 1726, Ap
	11	50.8	4.2	1362	9	US-09-938-842A-602	Sequence 602, App
	12	50.2	4.2	251	7	US-08-781-986A-3124	Sequence 3124, Ap
	13	39.2	3.3	1242	9	US-10-081-051-108	Sequence 108, App
	14	39.2	3.3	4122	9	US-10-081-051-105	Sequence 105, App
	15	38.8	3.2	215	10	US-09-294-093B-4258	Sequence 4258, Ap
	16	38.8	3.2	1198	9	US-09-895-913A-283	Sequence 283, App
	17	36.6	3.0	168	10	US-09-974-300-459	Sequence 459, App
	18	36.6	3.0	472	9	US-09-918-995-17334	Sequence 17334, A
	19	35.8	3.0	840	10	US-09-974-300-346	Sequence 346, App
	20	35.6	3.0	1649	10	US-09-925-302-121	Sequence 121, App
c	21	35.6	3.0	2119	9	US-10-043-487-68	Sequence 68, Appl
c	22	35.6	3.0	2617	10	US-09-764-864-283	Sequence 283, App
	23	35.2	2.9	228	10	US-09-867-550-37	Sequence 37, Appl
c	24	35	2.9	536165	9	US-09-939-964-1	Sequence 1, Appli
	25	34.4	2.9	1056	10	US-09-815-242-7886	Sequence 7886, Ap
	26	34	2.8	720	9	US-10-184-644-170	Sequence 170, App
	27	34	2.8	720	9	US-10-184-634-170	Sequence 170, App
	28	33.8	2.8	1551	10	US-09-815-242-7212	Sequence 7212, Ap
c	29	33.8	2.8	8161	10	US-09-759-152-9	Sequence 9, Appli
c	30	33.8	2.8	8175	10	US-09-759-152-7	Sequence 7, Appli
	31	33	2.7	864	10	US-09-974-300-288	Sequence 288, App
c	32	32.4	2.7	1176	10	US-09-815-242-7814	Sequence 7814, Ap
c	33	31.8	2.6	409	9	US-10-184-644-134	Sequence 134, App
c	34	31.8	2.6	409	9	US-10-184-634-134	Sequence 134, App
	35	31.8	2.6	882	9	US-10-184-644-574	Sequence 574, App
	36	31.8	2.6	882	9	US-10-184-634-574	Sequence 574, App
	37	31.8	2.6	2712	10	US-09-905-983-55	Sequence 55, Appl
	38	31.6	2.6	642	9	US-10-123-155-370	Sequence 370, App
c	39	31.6	2.6	734	9	US-10-184-644-458	Sequence 458, App
c	40	31.6	2.6	734	9	US-10-184-634-458	Sequence 458, App
	41	31.6	2.6	1032	9	US-10-184-644-552	Sequence 552, App
c	42	31.6	2.6	1032	9	US-10-184-644-552	Sequence 552, App
	43	31.6	2.6	1032	9	US-10-184-634-552	Sequence 552, App
c	44	31.6	2.6	1032	9	US-10-184-634-552	Sequence 552, App
	45	31.4	2.6	595	9	US-09-814-357-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-09-974-300-4739
; Sequence 4739, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4739
; LENGTH: 1217
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1217)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4739

Query Match 19.5%; Score 234.6; DB 10; Length 1217;
Best Local Similarity 51.3%; Pred. No. 5e-69;
Matches 593; Conservative 0; Mismatches 560; Indels 3; Gaps 2;

```
Qy      51 GGATGCGGGCGTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGA 110
           | || |    ||||||||||||||||||||| ||  |||  || ||       ||  ||||
Db      57 GAATCCTCTTGTCTATCTCGACAGCGCCGCCACTTCGCAAAAGCCAATCCAAGTCATTGA 116

Qy     111 AGCCACCCAACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGA 170
           |     | |  | | |||| | |     |       || ||||||||  ||  |  | |
Db     117 AAAGCTCGATGATTATTACCGCCGTTACAATTCTAATGTCCATCGTGGCGTGCATACGCT 176

Qy     171 AGCCCAACGCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAA 230
           |     |    ||| |     |||||||  ||||| || |||||   |   ||  ||||
```

```
Db    177 TGGTACGCTTGCGACTGATGAATATGAAGGCGCACGCGAAAAAGTCCGGGCTTTCTTGAA 236

Qy    231 TGCACCGGATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGT 290

Db    237 TGCGCAAGAAACGGCGGAGATTGTCTTTACTCGTGGCACAACCACGGCCATTAATTTAGT 296

Qy    291 GGCACAATGCTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGC 350

Db    297 GGCGCAAAGTTATGGAAATGACCATGTCGGCGAAGGCGATGAAATTGTCATTACGCCAAT 356

Qy    351 AGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGT 410

Db    357 GGAACATCATTCGAATATCATTCCATGGCAGCAGCTTGCCAAGCGGAAAGGGGCCGTGTT 416

Qy    411 GGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGAT 470

Db    417 GAAGTATATCCCTTTGCAAGAAGATGGGACGATTTCATTAGCTGATGTCGAAACAACGAT 476

Qy    471 TACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCC 530

Db    477 AACGGAACGAACAAAAATCGTTTCCGTTGTCTACGTTTCAAATGTGCTCGGCACGGTTAA 536

Qy    531 GGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGG 590

Db    537 CCCAATTGCTGAAATGGCGAAAGTCGCCCATAAACATGGCGCGGTCATGGTCGTCGATGG 596

Qy    591 TGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGC 650

Db    597 CGCCCAAGCTGCTCCCCATGTCCACGTGGATGTCCAACAATTAGACTGTGACTTTTTTGC 656

Qy    651 TTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGA 710

Db    657 TTTCTCGGCACACAAAATGTGTGGCCCTTCGGGCATCGGCGCCTTATACGGCAAAAAAGC 716

Qy    711 ACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTT 770

Db    717 GTTGCTTGAAGCGATGGAGCCAATTGAATTTGGCGGGGAAATGATTGATTTTGTCGGTCT 776

Qy    771 TGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGC 830

Db    777 GCAAGATTCGACTTGGAAAGAGCTCCCATGGAAGTTTGAAGGGGGAACGCCAATCATTGC 836

Qy    831 TGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCTGGCAGA-TTACGATATCAACCAGG 889

Db    837 TGGCGCAATCGGGTTCGGTGCTGCGATTGATTTTGTCACCGATTTANGGTTTAATGAAAT 896

Qy    890 CCGAAAGCTGGAGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCG 949

Db    897 TGAACAGCACGAAAAAGAACTCGTTGACTATTCGTTTGAGCGTCTCGGCGATTTCAAAGA 956

Qy    950 GCTTTCGTTCATTCCGCTGCCAG--GATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGT 1007

Db    957 TTTGACGATCTTTGGGCCAAAAGACCGTGCAGGGGTGATCACATTCCAGCTTGGCGACGT 1016

Qy   1008 TCATCATAGCGATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCA 1067

Db   1017 GCACCCCCACGATGTAGCGACTGTCCTTGATGCTGAAGGAATTGCTGTTCGCGCAGGCCA 1076

Qy   1068 GCATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTT 1127

Db   1077 TCATTGTGCACAACCGTTAATGAAATGGCTTGACGTTGTCGCCACTGCTCGTGCCAGCTA 1136

Qy   1128 TGCGCCATATAATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCT 1187

Db   1137 TTATGTGTACAATACGAAAGAAGATATTGATGCTCTGGCAAACGGATTAGTCAAAGCAAA 1196

Qy   1188 GGAATTATTGGTGGAT 1203

Db   1197 GGAGTATTTTAGCGAT 1212


RESULT 2
US-09-974-300-4738
```

```
; Sequence 4738, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4738
;   LENGTH: 1224
;   TYPE: DNA
;   ORGANISM: Bacillus clausii
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(1224)
;   OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4738

  Query Match             19.5%;  Score 234.6;  DB 10;  Length 1224;
  Best Local Similarity   51.3%;  Pred. No. 5e-69;
  Matches  593;  Conservative    0;  Mismatches  560;  Indels    3;  Gaps    2;

Qy      51 GGATGCGGGCGTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGA 110

Db      57 GAATCCTCTTGTCTATCTCGACAGCGCCGCCACTTCGCAAAAGCCAATCCAAGTCATTGA 116

Qy     111 AGCCACCCAACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGA 170

Db     117 AAAGCTCGATGATTATTACCGCCGTTACAATTCTAATGTCCATCGTGGCGTGCATACGCT 176

Qy     171 AGCCCAACGCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAA 230

Db     177 TGGTACGCTTGCGACTGATGAATATGAAGGCGCACGCGAAAAAGTCCGGGCTTTCTTGAA 236

Qy     231 TGCACCGGATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGT 290

Db     237 TGCGCAAGAAACGGCGGAGATTGTCTTTACTCGTGGCACAACCACGGCCATTAATTTAGT 296

Qy     291 GGCACAATGCTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGC 350

Db     297 GGCGCAAAGTTATGGAAATGACCATGTCGGCGAAGGCGATGAAATTGTCATTACGCCAAT 356

Qy     351 AGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGT 410

Db     357 GGAACATCATTCGAATATCATTCCATGGCAGCAGCTTGCCAAGCGGAAAGGGGCCGTGTT 416

Qy     411 GGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGAT 470

Db     417 GAAGTATATCCCTTTGCAAGAAGATGGGACGATTTCATTAGCTGATGTCGAAACAACGAT 476

Qy     471 TACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCC 530

Db     477 AACGGAACGAACAAAAATCGTTTCCGTTGTCTACGTTTCAAATGTGCTCGGCACGGTTAA 536

Qy     531 GGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGG 590

Db     537 CCCAATTGCTGAAATGGCGAAAGTCGCCCATAAACATGGCGCGGTCATGGTCGTCGATGG 596

Qy     591 TGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGC 650

Db     597 CGCCCAAGCTGCTCCCCATGTCCACGTGGATGTCCAACAATTAGACTGTGACTTTTTTGC 656

Qy     651 TTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGA 710

Db     657 TTTCTCGGCACACAAAATGTGTGGCCCTTCGGGCATCGGCGCCTTATACGGCAAAAAAGC 716
```

```
Qy    711 ACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTT 770
          |||| || ||||||  |||          |||||  ||||| || |    ||  || |
Db    717 GTTGCTTGAAGCGATGGAGCCAATTGAATTTGGCGGGGAAATGATTGATTTTGTCGGTCT 776

Qy    771 TGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGC 830
          | |  |  ||     ||      || |||||  | ||||  ||||||||||   | ||
Db    777 GCAAGATTCGACTTGGAAAGAGCTCCCATGGAAGTTTGAAGGGGGAACGCCAATCATTGC 836

Qy    831 TGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCTGGCAGA-TTACGATATCAACCAGG 889
          ||| |  || || ||  | || ||| | || |   |  | || ||| | | | |   |
Db    837 TGGCGCAATCGGGTTCGGTGCTGCGATTGATTTTGTCACCGATTTANGGTTTAATGAAAT 896

Qy    890 CCGAAAGCTGGAGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCG 949
          | |||  ||     | ||    ||  |           |    |||||
Db    897 TGAACAGCACGAAAAAGAACTCGTTGACTATTCGTTTGAGCGTCTCGGCGATTTCAAAGA 956

Qy    950 GCTTTCGTTCATTCCGCTGCCAG--GATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGT 1007
          |  || || ||  ||    ||    | |  |  || |  | ||  |  |||  | |||
Db    957 TTTGACGATCTTTGGGCCAAAAGACCGTGCAGGGGTGATCACATTCCAGCTTGGCGACGT 1016

Qy   1008 TCATCATAGCGATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCA 1067
          || |    |||| | | |||  | || |  |    | || |||||  | || || || ||
Db   1017 GCACCCCCACGATGTAGCGACTGTCCTTGATGCTGAAGGAATTGCTGTTCGCGCAGGCCA 1076

Qy   1068 GCATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTT 1127
          ||||| || || ||| || ||  |    | | |||   || |||    || |||   |
Db   1077 TCATTGTGCACAACCGTTAATGAAATGGCTTGACGTTGTCGCCACTGCTCGTGCCAGCTA 1136

Qy   1128 TGCGCCATATAATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCT 1187
          |     || ||||| ||    ||| | ||||| ||||  || |   | | |    ||
Db   1137 TTATGTGTACAATACGAAAGAAGATATTGATGCTCTGGCAAACGGATTAGTCAAAGCAAA 1196

Qy   1188 GGAATTATTGGTGGAT 1203
          ||| |  ||    |||
Db   1197 GGAGTATTTTAGCGAT 1212
```

```
RESULT 3
US-09-070-927A-290
; Sequence 290, Application US/09070927A
; Patent No. US20020120116A1
;   GENERAL INFORMATION:
;        APPLICANT: Charles A. Kunsch
;                   Patrick J. Dillon
;                   Steven Barash
;        TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
;        NUMBER OF SEQUENCES: 982
;        CORRESPONDENCE ADDRESS:
;             ADDRESSEE: Human Genome Sciences, Inc.
;             STREET: 9410 Key West Avenue
;             CITY: Rockville
;             STATE: Maryland
;             COUNTRY: USA
;             ZIP: 20850
;        COMPUTER READABLE FORM:
;             MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;             COMPUTER: HP Vectra 486/33
;             OPERATING SYSTEM: MSDOS version 6.2
;             SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;             APPLICATION NUMBER: US/09/070,927A
;             FILING DATE: 04-May-2000
;             CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;             APPLICATION NUMBER: 60/046,655
;             FILING DATE: 1997-05-16
;             APPLICATION NUMBER: 60/044,031
;             FILING DATE: 1997-05-06
;             APPLICATION NUMBER: 60/066,009
;             FILING DATE: 1997-11-14
;        ATTORNEY/AGENT INFORMATION:
;             NAME: Kenley K. Hoover
;             REGISTRATION NUMBER: 40,302
;             REFERENCE/DOCKET NUMBER: PB369
;        TELECOMMUNICATION INFORMATION:
;             TELEPHONE: (301) 309-8504
;             TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 290:
;        SEQUENCE CHARACTERISTICS:
;             LENGTH: 10397 base pairs
;             TYPE: nucleic acid
;             STRANDEDNESS: double
;             TOPOLOGY: linear
;        SEQUENCE DESCRIPTION: SEQ ID NO: 290:
US-09-070-927A-290

  Query Match             16.5%;  Score 198.4;  DB 10;  Length 10397;
  Best Local Similarity   50.0%;  Pred. No. 2.8e-56;
  Matches  583;  Conservative    0;  Mismatches  571;  Indels   12;  Gaps    3;

Qy     50 AGGATGCGGGCGTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTG 109
          | ||  ||   || ||| | ||||  || |||||  | |  |||||    || ||| | |
Db   5735 ATGAACCGCTAGTTTATTTAGACAATGCTGCGACAACCCAAAAACCAACGGCAGTGTTAG 5794

Qy    110 AAGCCACCCAACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCG 169
          | |     | ||| | |||              |  || || |||||  |    |    |
Db   5795 ATGTTTTACGACATTATTATGAAACCGATAATGCCAATGTTCATCGTGGGGTGCACACGT 5854

Qy    170 AAGCCCAACGCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGA 229
          ||||  ||||    ||   |  ||||||||  | || || || ||    |||||  |
Db   5855 TAGCCGAACGAGCAACGAAGGACTATGAAGCCTCCCGCGAAAAGGTTCGCCAATTTATCC 5914

Qy    230 ATGCACCGGATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGG 289
          ||||    ||       |    |  | |  || || || || ||          | ||   |
Db   5915 ATGCCAAAGAAACAGCAGAAGTTTTATTTACACGAGGAACAACAACCAGTTTGAATTGGA 5974

Qy    290 TGGCACAATGCTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGG 349
          | ||  || | ||||        |    |   | | || ||||| ||  |  |
Db   5975 TTGCCAAAAGTTATGGTGACTTGGCAGTTACAGCCGGGGATGAAATCGTCATATCTTACA 6034

Qy    350 CAGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAG 409
            ||||| ||  | ||| |  | |||||||      | || ||||  || || ||||
Db   6035 TGGAACATCATTCAAACATTATTCCCTGGCAACAATTAGCTCAACGTACAGGGGCCATTT 6094

Qy    410 TGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGA 469
          ||  | |  |     | | ||   |  | |   ||| | |         |   ||| |
Db   6095 TGAAGTATATAGACGTCACTGAAGACGGCTTTTTAGATATGGCAAGTGCACGTCAACAAA 6154

Qy    470 TTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCC 529
          ||||   |    ||| |  || | | ||| | || || ||    || |
Db   6155 TTACAGAAAAAACGAAAATTGTTTCGATTGCTCATGTTTCAAATGTGTTAGGGGTCATTA 6214

Qy    530 CGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATG 589
          | |   |   |    || |||||  |   ||   ||| || |||| || |
Db   6215 ATCCAATCGAAGAATTAACACAATTAGCTCATCAAAATGGCGCTGTGATGGTGGTAGACG 6274

Qy    590 GTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATG 649
          | || || |  |     ||| | || | ||| |||||      || |||  |||||| ||||
Db   6275 GCGCGCAAGCAGTTCCTCATATGCCTGTGGACGTTCAAGCCATCGATGCTGATTTTTATG 6334

Qy    650 CTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAG 709
          | |||   |||||||| ||| ||| || || || || || || || ||||| |||
Db   6335 CATTTAGTGGTCACAAAATGTGTGGACCAACTGGAATTGGTGTCCTTTATGGCAAACGTC 6394

Qy    710 AACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTT 769
          | |  ||||    |||   ||   |      || ||| ||||| || ||   ||    |
Db   6395 ATTTATTGGAACAAATGGAACCTGTGGAATTTGGTGGCGAAATGATTGACTTTGTTCATC 6454

Qy    770 TTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCG 829
          || | |     ||     ||      || |||||| | |||||||| || || ||| | |
Db   6455 TTCAAGAAAGTACTTGGAAAGAGCTTCCTTGGAAATTTGAAGCTGGCACACCTAATATTG 6514

Qy    830 CTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATAT-----CAA 884
          | || |  || |  ||| | || ||  | || |   ||||    | | |     | |
Db   6515 CCGGGGCAATTGCCTTAGGTGCCGCCATTGATTATTTAACAGAAATTGGTTTAGAGGCCA 6574
```

```
Qy    885 CCAGGCCGAAAGCTGGAGCCGTAGCTTAGCAACGCTGGCGGAAGATGC--GCTGGCGAAA 942
Db   6575 TTCATCAACATGAAGCAGCCCTTGTTCACTATGTTTTGCCGAAGTTGCAAGCTATCGAGG 6634

Qy    943 CGTCCCGGCTTTCGTTCATTCCGCTGCCAGGATTCCA-----GCCTGCTGGCCTTTGATT 997
Db   6635 GCTTGACGATTTATGGTCCACAGGATCCAAAAGACCACACAGGTGTGATTGCTTTTAATA 6694

Qy    998 TTGCTGGCGTTCATCATAGCGATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGC 1057
Db   6695 TTGAAGGTTTACATCCACATGATGTGGCTACGGCGTTAGATATGGAAGGGGTCGCTGTCA 6754

Qy   1058 GGGCCGGGCAGCATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGC 1117
Db   6755 GAGCGGGTCACCATTGCGCGCAACCGTTATTAAACTATTTGAGCGTGCCAGCAACCGCAC 6814

Qy   1118 GCGCCTCTTTTGCGCCATATAATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTG 1177
Db   6815 GGGCAAGTTTTTATTTATACAATACCAAAGAAGACGCAGATCGCTTAGTTGAAGCGATTA 6874

Qy   1178 ACCGCGCGCTGGAATTATTGGTGGAT 1203
Db   6875 AAGCGACAAAGGAGTTTTTCCAGCAT 6900
```

RESULT 4
US-08-781-986A-81/c
; Sequence 81, Application US/08781986A
; Publication No. US20030054436A1
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781,986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 81:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4280 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-81

Query Match 14.5%; Score 175.2; DB 7; Length 4280;
Best Local Similarity 47.7%; Pred. No. 1.4e-48;
Matches 550; Conservative 0; Mismatches 593; Indels 9; Gaps 1;

```
Qy     64 TATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAACAG 123
Db   3337 TATCTTGATTCAACAGCGACAAGTCAAACGCCTGTGCAAGTGTTAAATGTTTTAGAAGAT 3278

Qy    124 TTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGCCTG 183
Db   3277 TACTACAAGCGTTATAATTCAAACGTTCATCGTGGTGTTCATACATTAGGATCATTGGCA 3218

Qy    184 ACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGAT 243
Db   3217 ACTGATGGTTATGAAAATGCCCGTGAAACCGTTCGTCGTTTTATTAATGCGAAGTATTTT 3158

Qy    244 AAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTAT 303
Db   3157 GAAGAAATCATTTTTACACGCGGAACAACTGCGTCGATTAACCTTGTAGCACATAGCTAT 3098

Qy    304 GCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCC 363
Db   3097 GGTGATGCAAATGTTGAAGAGGGCGATGAAATTGTTGTCACTGAAATGGAACATCATGCC 3038

Qy    364 AACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCG 423
Db   3037 AATATTGTTCCTTGGCAACAGTTAGCAAAGCGTAAAAATGCGACATTGAAATTTATACCA 2978

Qy    424 CTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGT 483
Db   2977 ATGACAGCTGACGGTGAATTAAACATCGAGGATATTAAGCAAACGATTAATGATAAAACA 2918

Qy    484 CGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGA 543
Db   2917 AAGATCGTTGCTATTGCACATATATCTAATGTGCTCGGTACAATTAATGATGTTAAAACC 2858

Qy    544 GCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCA 603
Db   2857 ATTGCAGAAATAGCTCATCAACATGGTGCAATTATCAGTGTTGATGGGGCGCAAGCAGCA 2798

Qy    604 GTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCAC 663
Db   2797 CCACATATGAAACTTGATATGCAAGAAATGAATGCTGATTTTTATAGTTTTAGTGGTCAT 2738

Qy    664 AAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCG 723
Db   2737 AAAATGCTTGGACCAACAGGTATTGGCGTATTATTTGGTAAACGTGAGTTACTACAAAAA 2678

Qy    724 ATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACG 783
Db   2677 ATGGAACCGATTGAGTTCGGTGGCGACATGATTGATTTTGTAAGTAAGTATGATGCAACA 2618

Qy    784 ACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGA 843
Db   2617 TGGGCTGATTTACCTACTAAATTTGAGGCGGGTACTCCATTAATTGCTCAAGCAATTGGG 2558

Qy    844 TTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGGAGC 903
Db   2557 CTTGCAGAAGCTATTCGCTATTTAGAACGCATAGGTTTTGATGCAATTCATAAATATGAA 2498

Qy    904 CGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCATTC 963
Db   2497 CAAGAATTAACGATATATGCTTATGAGCAAATGTCTGCAATTGAAGGAATTGAAATTTAT 2438

Qy    964 CGCTGCCAGGA---------TTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCAT 1014
Db   2437 GGCCCGCCAAAGGATCGTCGTGCAGGTGTAATAACGTTTAATTTACAAGATGTACATCCA 2378

Qy   1015 AGCGATATGGTGACGCTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGC 1074
Db   2377 CACGATGTTGCTACAGCCGTAGATACAGAAGGTGTAGCGGTTAGAGCTGGGCATCATTGT 2318

Qy   1075 GCTCAGCCGCTACTGGCAGAATTAGGCGTAACCGGCACACTGCGCCGCCTCTTTTGCGCCA 1134
Db   2317 GCGCAACCGTTAATGAAATGGTTAAATGTGTCTTCAACAGCTAGAGCGAGTTTTTATATA 2258

Qy   1135 TATAATACAAAGAGTGATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTA 1194
```

```
Db   2257 TACAACACGAAAGAAGACGTTGATCAGTTAATAAATGCCTTGAAACAAACGAAGGAGTTT 2198

Qy   1195 TTGGTGGATTAA 1206
          ||     || ||
Db   2197 TTCTCTTATGAA 2186
```

RESULT 5
US-09-974-300-304
; Sequence 304, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression'.
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 304
; LENGTH: 727
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-304

Query Match 12.0%; Score 145.2; DB 10; Length 727;
Best Local Similarity 51.5%; Pred. No. 8.8e-39;
Matches 333; Conservative 0; Mismatches 313; Indels 0; Gaps 0;

```
Qy     60 CGTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCA 119
          ||| ||| | |||||||| || ||| ||| || ||    | |||  ||||  |     |
Db     66 CGTTTATTTGGACAGCGCGGCCACCTCGCAAAAGCCGCGCGTCGTCATTGAGACGCTTGA 125

Qy    120 ACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACG 179
            ||| |||||       |      || |||||||| ||     | |   ||    |
Db    126 TGAGTATTACAACTCTTACAATTCCAATGTCCATCGCGGCGTCCATACATTAGGAACAAA 185

Qy    180 CCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGA 239
             || |   | |||||||  ||||| |||||  ||   | | ||  | || ||
Db    186 AGCAACGGATGGATATGAAGGAGCACGTGAGAAGGTCAGGAAGTTTATAAACGCGAAATC 245

Qy    240 TGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATG 299
             |   ||  | |  ||  | || || ||     || | |||| ||| || |   |
Db    246 GATGCAGGAAATTATTTTCACAAGGGGTACGACAACGGCCCTGAACACGGTTGCGCTCAG 305

Qy    300 CTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCA 359
          |||||| ||  |    ||  | || || ||||| ||  |  | |    |   || |||||
Db    306 CTATGCCCGCGCAAACCTTAAGCCCGGTGATGAAATCGTCATTACACCGATGGAGCACCA 365

Qy    360 CGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATT 419
           || ||  |  | || |||| |  ||  | | ||  ||| || |||   ||  | |  |
Db    366 TGCGAATATTATTCCGTGGCAGCAGGCTGTCAAAGCAACCGGCGCCACGCTGAAGTACAT 425

Qy    420 GCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCG 479
          |||||| | |   |   |          | ||    |   |||  | | ||
Db    426 TCCGCTTCAGGAAGACGGCACCATTTCCTTGGACGATGTCAGGGAAACGGTGACGGAAAA 485

Qy    480 TAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGC 539
          ||     ||  |  |    |  ||  | || ||||||   ||       |       |
Db    486 TACAAAAATCGTCTCTGCAGCGCATGTTTCAAACGTTCTCGGTACGATCAACCCGATTAA 545

Qy    540 GCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGG 599
          |  |  |   |  | || ||   |  |||   |||  || | || || ||||| ||  |
Db    546 GGAAATGGCCAAAATCGCCCACGAACACGGCGCGGTCATGGTCGTAGACGGTGCGCAAAG 605

Qy    600 GGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGG 659
            |  |||  |       || || ||| |  ||||    |||||||||  | |||||||
```

```
Db    606 CACGCCGCACATGAAAATTGACGTCCAGGATTTGGACTGCGATTTCTATACGTTTTCAGC 665

Qy    660 TCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAA 705
           |||||| ||| ||| ||||| || ||||| ||  |||| | |||
Db    666 CCACAAAATGTGTGGTCCGACCGGAATCGGGGTATTGTAGGAAAAA 711
```

RESULT 6
US-09-815-242-3775/c
; Sequence 3775, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3775
; LENGTH: 265
; TYPE: DNA
; ORGANISM: Salmonella typhimurium
US-09-815-242-3775

Query Match 11.4%; Score 138; DB 10; Length 265;
Best Local Similarity 80.2%; Pred. No. 1.5e-36;
Matches 162; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

```
Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||| |||||||||| |||||||||||| |||||||||| |||   ||||||||
Db    202 ATGAACGCTTTTAATCCCACGCAGTTTCGCGCGCAGTTTCCCGCGCTAGCCGATGCGGGT 143

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          || |||||||| |||||||| || ||  | || ||  | || ||  |||| ||||| ||
Db    142 GTTTATCTCGATAGCGCCGCCACGGCATTAAAGCCACAGGCAGTCATTGACGCCACGCAC 83

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||  || |||||||||| ||||| ||||| ||||||||||  | || || |||
Db     82 CAGTTTTATTGTTTGAGCGCCGGTAACGTTCATCGTAGCCAGTTTGCGCAGGCGCAGCGC 23

Qy    181 CTGACCGCGCGTTATGAAGCTG 202
          ||||| ||||  |||||||| |
Db     22 CTGACGGCGCAATATGAAGCGG 1
```

RESULT 7
US-09-738-626-1727
; Sequence 1727, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:

```
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 09/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1727
;   LENGTH: 1263
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1727

  Query Match             9.9%;  Score 119.4;  DB 9;  Length 1263;
  Best Local Similarity   46.1%;  Pred. No. 6.7e-30;
  Matches  543;  Conservative    0;  Mismatches  606;  Indels   30;  Gaps    3;

Qy      51 GGATGCGGGCGTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGA 110
           | |  ||   |  || || |||   | | ||||  |||   ||| ||   |||
Db      90 GAAACCGCTTGCTTACCTGGACTCAGGTGCGACATCGCAGCGACCCGAGCGGGTGTGGCG 149

Qy     111 AGCCACCCAACAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGA 170
           ||     | || |||      |  | |  ||    ||| || ||| |    |
Db     150 TGCAGAGGAGCACTTTGTGCTGCACACCAACGCCCCCGTGCACCGCGGTGCCTACCAACT 209

Qy     171 AGCCCAACGCCTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAA 230
           ||  |       || |    |||||||| ||| || |||||  | ||    ||  |
Db     210 GGCTGAGGAAGCAACGGATGCTTATGAAGGTGCCCGCGAGAAGATCGCTGCCTTTGTTGG 269

Qy     231 TGCACCGGATGATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCA---------- 280
           |||   | |  || ||| |    ||   | |   |  ||    | |  ||
Db     270 TGCCGAGCAGCATGAAATTGCGTTCACTAAGAATGCAACTGAAGCACTCAATCTTGTTGC 329

Qy     281 --TCAACATGGTGGCACAATGCTATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTAT 338
             || | |||  |   |  | |  |        ||| | ||  | || |||    |  |
Db     330 GTACACCTTGGGTGATGACCGTTCCGGTAAGTATCGTGTCCAGGCCGGGGATACCGTGGT 389

Qy     339 TGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAAC 398
             |||  | |  ||| |||||||| ||| | || || |||| |  | |   ||  | |||
Db     390 CATCACGGAGCTAGAGCACCACGCAAACTTGGTGCCATGGCAGGAGCTGTGCCGTCGAAC 449

Qy     399 TGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTT 458
           || || | | ||  |   |     | | | ||   |  | |     ||| ||||||
Db     450 CGGTGCGACATTGAAGTGGTACAAGGTGACTGAAGATGGTCGCATTGATCTCGATT---- 505

Qy     459 GCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTAC 518
             ||     | ||  |      |     | |  | || ||   |||   ||| || || ||
Db     506 --CACTCGAGCTTGATGAAACTGTCAAGGTCGTTGCCTTCACTCACCAGTCCAATGTGAC 563

Qy     519 TGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGT 578
             || | |    | ||| |  |      | ||     |||  |  | | |||     |
Db     564 CGGTGCTGTGGCTGATGTTCCAGAGTTGGTTCGTCGTGCCAAGGCTGTCGGCGCTCTCAC 623

Qy     579 GATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATAT 638
           | || ||||||    |||  |     ||| | || | | || | ||  | ||||||  |
Db     624 GGTGCTTGATGCGTGCCAGTCTGTTCCTCATATGCCAGTGAATTTCCACGAGCTGGATGT 683

Qy     639 TGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTA 698
```

```
           ||||||| ||| || || || || ||  ||   || ||  ||||  | |||||  ||||
Db     684 AGATTTCTCTGCATTCTCTGGCCATAAGATGCTGGGACCTGCAGGCGTGGGCGTTGTGTA 743

Qy     699 TGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCA 758
           ||  || ||   | |  |||| |   || | || |   ||  || ||    ||| || |
Db     744 TGCAAAGTCCCCAATCTTGGATGAACTGCCACCATTTTTGACTGGTGGTTCCATGATTGA 803

Qy     759 CGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAAC 818
           |  || |   | || || | |||         | || ||       | || || || ||
Db     804 AGTTGTCACCATGGAGGGTTCCACCTACGCTGCCGCACCTCAACGTTTTGAGGCCGGCAC 863

Qy     819 GCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCTGGCAGATTACGA 878
           ||  |     |       ||  | || ||  | || ||  ||||   ||||   ||   ||
Db     864 GCAGATGACCAGCCAGGTTGTGGGCTTGGGTGCTGCCGTGGACATGCTGAATGAAATCGG 923

Qy     879 TATCAACCAGGCCGAAAGCTGGAGCCGTAGCTTAGCAACGCTGGCGGAAGATGCGCTGGC 938
           |||  |       || |          |      ||  |  |  |||   ||   |||  |
Db     924 TATGGAAGCAATCGCAGCGCATGAGCACGCATTGACTGCTTACGCGTTGGAAAAGCTCAC 983

Qy     939 GAAACGTCCCGGCTT----------TCGTTCATTCCGCTGCCAGGATTCCAGCCTGCT 986
           |  |  |   ||  |          || ||         |||     | |     |
Db     984 GGCAATTAAGGGACTAACCATTGCTGGTCCTTTGACTGCAGAGCAGCGCGGCGGTGCAAT 1043

Qy     987 GGCCTTTGATTTTGCTGGCGTTCATCATAGCGATATGGTGACGCTGCTGGCGGAGTACGG 1046
             ||| | | | |  ||| |||| |    |||| | |     |||| |  ||  | ||
Db    1044 CAGCTTCGGTGTCGAGGGCATTCACCCACACGATCTAGGCCAAGTGCTTGACGATCAGGG 1103

Qy    1047 TATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGTAAC 1106
             |    | || | ||| || || |||||   |||  | |        |   ||||
Db    1104 CGTGAATATCCGCGTCGGCCACCACTGCGCGTGGCCCGTGCACCGCAGCATGAACGTACA 1163

Qy    1107 CGGCACACTGCGCGCCTCTTTTGCGCCATATAATACAAAGAGTGATGTGGATGCGCTGGT 1166
             |||    | || |||||    |  |.||| ||      ||  | ||    || |
Db    1164 ATCGACAGCAAGAGCATCTTTCTATCTCTATAACACCTTCGAAGAAATCGACCGCCTCGC 1223

Qy    1167 GAATGCCGTTGACCGCGCGCTGGAATTATTGGTGGATTA 1205
           |  ||  | ||    ||   | |||| || |  | | |
Db    1224 GGCAGCGATCGAGAAGGCCAAGCAATTCTTTGGAGTTGA 1262


RESULT 8
US-09-878-574-3815
; Sequence 3815, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
;  APPLICANT: Byrum, Joseph R.
;  APPLICANT:  La Rosa, Thomas J.
;  APPLICANT:  Thompson, Michael D.
;  TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
;  TITLE OF INVENTION:  Plants
;  FILE REFERENCE: 38-21(15401)B
;  CURRENT APPLICATION NUMBER: US/09/878,574
;  CURRENT FILING DATE:  2001-12-21
;  PRIOR APPLICATION NUMBER: 09/333,535
;  PRIOR FILING DATE: 1999-06-14
;  NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 3815
;   LENGTH: 419
;   TYPE: DNA
;   ORGANISM: Glycine max
;   OTHER INFORMATION: Clone ID: LIB3028-005-Q1-B1-D7
US-09-878-574-3815

  Query Match             6.0%;  Score 72.6;  DB 10;  Length 419;
  Best Local Similarity   54.2%;  Pred. No. 3e-14;
  Matches  147;  Conservative    0;  Mismatches  124;  Indels    0;  Gaps    0;

Qy     576 GGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGA 635
           | || | || ||||     || | |      ||  |       ||||| |||   ||  |
Db     149 GTTGCTTGTAGATGCATGCCACAGTGTTCCACACATGATGATTGATGTCCAGAGCCTTAA 208

Qy     636 TATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCT 695
           | |||| ||   ||    |||   ||||||  ||| ||||||  ||||| || ||  | ||
```

```
Db    209 TGTTGACTTTCTTGTGGCTTCTTCTCACAAGATGTGTGGCCCTACAGGCATTGGATTCCT 268

Qy    696 GTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCCTGGCTGGGCGGCGGCAAAATGGT 755
          || ||||||  | | ||  ||    | ||| | || |   ||||  | ||  ||||| |
Db    269 ATACGGTAAAATATACCTCTTGTCTTCCATGCCTCCATTTTTGGGTTGTGGTGAAATGAT 328

Qy    756 TCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGG 815
          |  ||| ||   | ||||    |  ||   |       | ||    | | | ||| ||||
Db    329 TTCCGATGTATATCTTGATCATTCAACTTATTCCGAACCTCCAACCATATTTGAACCTGG 388

Qy    816 AACGCCAAATGTCGCTGGTGTCATAGGATTA 846
          ||| |||  |   |  |  |  || || |||
Db    389 AACACCAGCTAGTGGGGAAGCAATTGGTTTA 419


RESULT 9
US-09-974-300-449
; Sequence 449, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
;  APPLICANT: Berka, Randy M.
;  APPLICANT:  Clausen, Ib Groth
;  TITLE OF INVENTION: Methods For Monitoring Multiple Gene
;  TITLE OF INVENTION:  Expression
;  FILE REFERENCE: 10085.500-US
;  CURRENT APPLICATION NUMBER: US/09/974,300
;  CURRENT FILING DATE:  2001-10-05
;  PRIOR APPLICATION NUMBER: 09/680,598
;  PRIOR FILING DATE: 2000-10-06
;  PRIOR APPLICATION NUMBER: 60/279,526
;  PRIOR FILING DATE: 2001-03-27
;  NUMBER OF SEQ ID NOS: 8481
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 449
;   LENGTH: 246
;   TYPE: DNA
;   ORGANISM: Bacillus licheniformis
US-09-974-300-449

  Query Match             4.9%;  Score 58.8;  DB 10;  Length 246;
  Best Local Similarity   54.1%;  Pred. No. 1.1e-09;
  Matches  120;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy     975 TTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGATATGGTGACGCTGCT 1034
           | |  | ||  || | ||  |  |||  | ||||||||    |||| |||  ||  ||||
Db      23 TGCAGGACTCGTGACGTTCAACCTTGAAGACGTTCATCCGCACGATGTGGCAACCGTGCT 82

Qy    1035 GGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAGCCGCTACTGGCAGA 1094
           ||  |   | ||  | ||| |  | ||||| || || ||||| |||||  || ||  |
Db      83 GGATGCAGAAGGGGTCGCCGTCAGAGCCGGCCATCACTGCGCGCAGCCTTTAATGAAATG 142

Qy    1095 ATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAATACAAAGAGTGATGT 1154
             |   |||  ||| |||     | ||    |||   |  ||||||||  |    ||  |
Db     143 GCTGAACGTCTCCGCCACTGCAAGAGCGAGCTTTTATCTGTATAATACGGAAGAAGACAT 202

Qy    1155 GGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATT 1196
            ||    |||  |  | ||  || |      |   ||| |  ||
Db     203 CGACAAACTGATTGAAGCTCTTCAGAAGACAAAGGAGTATTT 244


RESULT 10
US-09-738-626-1726/c
; Sequence 1726, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT:  MIZOGUCHI, HIROSHI
;  APPLICANT:  ANDO, SEIKO
;  APPLICANT:  HAYASHI, MIKIRO
;  APPLICANT:  OCHIAI, KEIKO
;  APPLICANT:  YOKOI, HARUHIKO
;  APPLICANT:  TATEISHI, NAOKO
;  APPLICANT:  SENOH, AKIHIRO
;  APPLICANT:  IKEDA, MASATO
;  APPLICANT:  OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE:  2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
;  PRIOR APPLICATION NUMBER: JP 00/159162
;  PRIOR FILING DATE: 2000-04-07
;  PRIOR APPLICATION NUMBER: JP 00/280988
;  PRIOR FILING DATE: 2000-08-03
;  NUMBER OF SEQ ID NOS: 7059
;  SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 1726
;   LENGTH: 162
;   TYPE: DNA
;   ORGANISM: Corynebacterium glutamicum
US-09-738-626-1726

  Query Match             4.6%;  Score 55.6;  DB 9;  Length 162;
  Best Local Similarity   59.5%;  Pred. No. 1.1e-08;
  Matches   94;  Conservative    0;  Mismatches   64;  Indels    0;  Gaps    0;

Qy     609 TTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACT 668
           | | || | | || | ||  | |||||| | ||||||| ||| || || || || ||  |
Db     162 TATGCCAGTGAATTTCCACGAGCTGGATGTAGATTTCTCTGCATTCTCTGGCCATAAGAT 103

Qy     669 GTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTC 728
           |   || ||  ||||  | |||||  ||||||  || ||   | |  |||| |   || |
Db     102 GCTGGGACCTGCAGGCGTGGGCGTTGTGTATGCAAAGTCCCCAATCTTGGATGAACTGCC 43

Qy     729 GCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGA 766
             || |   ||   || ||    ||| || | |  || |
Db      42 ACCATTTTGACTGGTGGTTCCATGATTGAAGTTGTCA 5


RESULT 11
US-09-938-842A-602
; Sequence 602, Application US/09938842A
; Patent No. US20020160378A1
; GENERAL INFORMATION:
;  APPLICANT: Harper, Jeff
;  APPLICANT:  Kreps, Joel
;  APPLICANT:  Wang, Xun
;  APPLICANT:  Zhu, Tong
;  TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAININ
;  TITLE OF INVENTION:  SAME, AND METHODS OF USE
;  FILE REFERENCE: SCRIP1300-3
;  CURRENT APPLICATION NUMBER: US/09/938,842A
;  CURRENT FILING DATE:  2001-08-24
;  PRIOR APPLICATION NUMBER: US 60/227,866
;  PRIOR FILING DATE: 2000-08-24
;  PRIOR APPLICATION NUMBER: US 60/264,647
;  PRIOR FILING DATE: 2001-01-16
;  PRIOR APPLICATION NUMBER: US 60/300,111
;  PRIOR FILING DATE: 2001-06-22
;  NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 602
;   LENGTH: 1362
;   TYPE: DNA
;   ORGANISM: Arabidopsis thaliana
US-09-938-842A-602

  Query Match             4.2%;  Score 50.8;  DB 9;  Length 1362;
  Best Local Similarity   47.7%;  Pred. No. 1.4e-06;
  Matches  148;  Conservative    0;  Mismatches  162;  Indels    0;  Gaps    0;

Qy     391 CAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATGTC 450
           || |||   |||    | ||     | || ||  | ||  |  |  || ||   ||| |
Db     487 CAGCAAGAAGGATTTGAGGTAACTTATTTACCTGTGAAAACTGATGGATTGGTTGATTTA 546

Qy     451 GATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGTCG 510
```

```
Db    547 GAGATGTTGAGAGAAGCTATTAGGCCAGACACAGGGCTAGTTTCTATTATGGCTGTGAAC 606

Qy    511 AACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCGGG 570

Db    607 AATGAGATTGGTGTGGTTCAACCTATGGAGGAGATTGGTATGATTTGCAAAGAGCATAAT 666

Qy    571 ATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAA 630

Db    667 GTTCCGTTTCATACTGATGCTGCTCAAGCTATTGGGAAGATACCTGTTGATGTTAAGAAG 726

Qy    631 CTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGC 690

Db    727 TGGAATGTTGCTTTGATGTCTATGAGTGCTCACAAGATCTATGGACCGAAAGGTGTTGGT 786

Qy    691 GTGCTGTATG 700

Db    787 GCTTTGTATG 796


RESULT 12
US-08-781-986A-3124
; Sequence 3124, Application US/08781986A
; Publication No. US20030054436A1
;  GENERAL INFORMATION:
;    APPLICANT:  Charles Kunsch
;    TITLE OF INVENTION:  Staphylococcus aureus Polynucleotides and Sequences
;    NUMBER OF SEQUENCES:  5255
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/781,986A
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Benson, Bob
;      REGISTRATION NUMBER:  30,446
;      REFERENCE/DOCKET NUMBER:  PB248PP
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  3124:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  251 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  double
;      TOPOLOGY:  linear
US-08-781-986A-3124

  Query Match             4.2%;  Score 50.2;  DB 7;  Length 251;
  Best Local Similarity   59.4%;  Pred. No. 9.5e-07;
  Matches   85;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy    626 AGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTA 685

Db      8 AGAAATGAACTGCTGATTTTTATAGTTTTAGTGGTCATAAAATGCTTGGACCAACAGGTA 67

Qy    686 TCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCG 745

Db     68 TTGGCGTATTATTTGGTAAACGTGAGTTACTACAAAAAATGGAACCGATTGAGTTCGGTG 127

Qy    746 GCAAAATGGTTCACGAAGTGAGT 768

Db    128 GCGACATGATTGATTTTGTAAGT 150


RESULT 13
US-10-081-051-108
; Sequence 108, Application US/10081051
; Publication No. US20030044422A1
; GENERAL INFORMATION:
;  APPLICANT: Barbet, Anthony F.
;  APPLICANT:  Whitmire, William M.
;  APPLICANT:  Kamper, Sondra M.
;  APPLICANT:  Simbi, Bigboy H.
;  APPLICANT:  Ganta, Roman R.
;  APPLICANT:  Moreland, Annie L.
;  APPLICANT:  Mwangi, Duncan M.
;  APPLICANT:  McGuire, Travis C.
;  APPLICANT:  Mahan, Suman M.
;  TITLE OF INVENTION: Ehrlichia Ruminantium Polypeptides, Antigens, Polynucleotides
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: UF-299XC1
;  CURRENT APPLICATION NUMBER: US/10/081,051
;  CURRENT FILING DATE:  2002-02-20
;  PRIOR APPLICATION NUMBER: US 60/269,944
;  PRIOR FILING DATE: 2001-02-20
;  NUMBER OF SEQ ID NOS: 117
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 108
;  LENGTH: 1242
;  TYPE: DNA
;  ORGANISM: Ehrlichia ruminantium (formerly Cowdria ruminantium)
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(1242)
;  OTHER INFORMATION: Corresponds to SEQ ID NO:105, nucleotides 2099..3340
;  OTHER INFORMATION: Hypothetical iron-sulfur co-factor synthesis
;  OTHER INFORMATION: Product = "11hworf2"
US-10-081-051-108

  Query Match             3.3%;  Score 39.2;  DB 9;  Length 1242;
  Best Local Similarity   58.6%;  Pred. No. 0.011;
  Matches   68;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy    585 TGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTT 644

Db    558 TGATGCAGCTCAAGCATTTGGTAAAATACCAATAGATGTAAACAAAATGAATATAGATTT 617

Qy    645 CTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATG 700

Db    618 GCTTAGTATATCAGGACATAAAATATATGCTCCAATGGGAATAGGGGCATTATATG 673


RESULT 14
US-10-081-051-105
; Sequence 105, Application US/10081051
; Publication No. US20030044422A1
; GENERAL INFORMATION:
;  APPLICANT: Barbet, Anthony F.
;  APPLICANT:  Whitmire, William M.
;  APPLICANT:  Kamper, Sondra M.
;  APPLICANT:  Simbi, Bigboy H.
;  APPLICANT:  Ganta, Roman R.
;  APPLICANT:  Moreland, Annie L.
;  APPLICANT:  Mwangi, Duncan M.
;  APPLICANT:  McGuire, Travis C.
;  APPLICANT:  Mahan, Suman M.
;  TITLE OF INVENTION: Ehrlichia Ruminantium Polypeptides, Antigens, Polynucleotides
;  TITLE OF INVENTION:  Methods of Use
;  FILE REFERENCE: UF-299XC1
;  CURRENT APPLICATION NUMBER: US/10/081,051
;  CURRENT FILING DATE:  2002-02-20
```

```
;  PRIOR APPLICATION NUMBER: US 60/269,944
;  PRIOR FILING DATE: 2001-02-20
;  NUMBER OF SEQ ID NOS: 117
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 105
;   LENGTH: 4122
;   TYPE: DNA
;   ORGANISM: Ehrlichia ruminantium (formerly Cowdria ruminantium)
US-10-081-051-105

  Query Match             3.3%;  Score 39.2;  DB 9;  Length 4122;
  Best Local Similarity   58.6%;  Pred. No. 0.02;
  Matches   68;  Conservative    0;  Mismatches   48;  Indels    0;  Gaps    0;

Qy     585 TGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTT 644
           ||||| ||||| |     |   |.| || |||||  |  || || |||| |||||
Db    2656 TGATGCAGCTCAAGCATTTGGTAAAATACCAATAGATGTAAACAAAATGAATATAGATTT 2715

Qy     645 CTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATG 700
             |  | | ||||| || ||| | ||||  || |  || || || |   | ||||
Db    2716 GCTTAGTATATCAGGACATAAAATATATGCTCCAATGGGAATAGGGGCATTATATG 2771


RESULT 15
US-09-294-093B-4258
; Sequence 4258, Application US/09294093B
; Patent No. US20010051335A1
; GENERAL INFORMATION:
;  APPLICANT:      Lalgudi, Raghunath, V.
;  APPLICANT:  Ito, Laura, Y.
;  APPLICANT:  Sherman, Bradley, K.
;  TITLE OF INVENTION:  POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
;  FILE REFERENCE:  PL-0009 US
;  CURRENT APPLICATION NUMBER: US/09/294,093B
;  CURRENT FILING DATE:  1999-04-16
;  PRIOR APPLICATION NUMBER:  60/082,567
;  PRIOR FILING DATE:  April 21, 1998
;  NUMBER OF SEQ ID NOS:      6207
;  SOFTWARE:       PERL Program
; SEQ ID NO 4258
;   LENGTH: 215
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20010051335A1 700353591H1
;   NAME/KEY: unsure
;   LOCATION: 5
;   OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-4258

  Query Match             3.2%;  Score 38.8;  DB 10;  Length 215;
  Best Local Similarity   58.8%;  Pred. No. 0.0065;
  Matches   67;  Conservative    0;  Mismatches   47;  Indels    0;  Gaps    0;

Qy     585 TGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTT 644
           ||||| ||| |||| |   | | |  ||||    |||||  |     |||  ||||   |
Db      15 TGATGCTGCGCAGGCGCTGGGGAAGATCCCGATTGATGTGAACAGGATGGGGATTGGGCT 74

Qy     645 CTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTA 698
           |    | || || || |||||  | |||||||| |  ||  | || |  || ||
Db      75 CATGTCCTTGTCCGGCCACAAGATTTATGGCCCCAAGGGCGTGGGGGCCCTCTA 128


Search completed: April 22, 2003, 15:13:58
Job time : 187.262 secs
```

GenCore version 5.1.4_p5_4578

OM nucleic - nucleic search, using sw model

Run on: April 22, 2003, 09:05:34 ; Search time 1966.07 Seconds
(without alignments)
9934.410 Million cell updates/sec

Title: US-09-622-419-3
Perfect score: 1206
Sequence: 1 atgaacgtttttaatcccgc.....tggaattattggtggattaa 1206

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	351	29.1	571	9	AF373211	AF373211 AF373211
	2	333.4	27.6	656	17	CNS01026	AL152975 Anopheles
c	3	320	26.5	932	17	AF095401	AF095401 AF095401
	4	138.4	11.5	755	14	BQ518982	BQ518982 EST626397
	5	136.8	11.3	725	13	BM407607	BM407607 EST581922
	6	135.4	11.2	645	14	BQ578320	BQ578320 WHE0301_E
	7	127.8	10.6	552	10	AV550366	AV550366 AV550366
	8	126.6	10.5	656	10	AV914366	AV914366 AV914366
	9	122.6	10.2	713	17	BH242226	BH242226 ATZFC20TF
	10	116.2	9.6	635	17	BH242467	BH242467 ATZFC60TR
c	11	115.2	9.6	707	17	BH374662	BH374662 AG-ND-179
	12	114.8	9.5	591	10	BE511577	BE511577 946062A01
	13	114.6	9.5	735	17	BH242443	BH242443 ATZFD83TF
c	14	114.6	9.5	950	17	AF035972	AF035972 AF035972
	15	111.6	9.3	669	17	BH242236	BH242236 ATZFD22TF
c	16	103.6	8.6	423	17	BH392425	BH392425 AG-ND-178
	17	103.6	8.6	684	10	AV917435	AV917435 AV917435
	18	102.6	8.5	596	10	AW459362	AW459362 sh41a11.y
	19	101	8.4	540	9	AJ470647	AJ470647 AJ470647
	20	98	8.1	417	17	BH242524	BH242524 ATZFC66TR
	21	97.4	8.1	542	9	AI773920	AI773920 EST255020
	22	94.8	7.9	660	14	BQ138885	BQ138885 NF008F12P
	23	92.4	7.7	461	10	AV908881	AV908881 AV908881
	24	87.8	7.3	666	10	AV822179	AV822179 AV822179
	25	86.2	7.1	551	10	AV432069	AV432069 AV432069
	26	85.6	7.1	639	10	AW689739	AW689739 NF023H01S
c	27	82.6	6.8	530	17	BH242301	BH242301 ATZFD81TR
	28	76.6	6.4	416	10	AV413701	AV413701 AV413701
c	29	76.2	6.3	487	17	BH242276	BH242276 ATZFD94TR
	30	75.8	6.3	803	11	AY104487	AY104487 Zea mays
	31	75.2	6.2	550	10	AW733805	AW733805 sk84e11.y
	32	75.2	6.2	675	14	BQ995735	BQ995735 QGG10N06.
	33	75	6.2	541	14	BM887196	BM887196 sam35d09.
	34	74.6	6.2	585	10	AW398793	AW398793 EST309293
	35	74.6	6.2	777	12	BG645380	BG645380 EST506999
	36	72.4	6.0	596	10	AW254947	AW254947 ML1117 pe
	37	71.8	6.0	676	12	BG450659	BG450659 NF096E03D
	38	71.2	5.9	342	17	BH242235	BH242235 ATZFB87TF
	39	69.4	5.8	569	12	BF484687	BF484687 WHE2318_E
	40	68.8	5.7	481	17	AZ049189	AZ049189 GSSBru059
	41	67	5.6	621	12	BG908875	BG908875 TaLr1171H
	42	64.4	5.3	358	10	AW696435	AW696435 NF104A01S
c	43	64.4	5.3	524	17	B96222	B96222 T23A3TRB TA
	44	63	5.2	258	13	BM442494	BM442494 EBan01_SQ
	45	62.6	5.2	314	17	AZ578065	AZ578065 18c10 Sho

ALIGNMENTS

RESULT 1
AF373211/c
LOCUS AF373211 571 bp mRNA linear EST 03-MAY-2002
DEFINITION AF373211 Gossypium hirsutum cotyledon Gossypium hirsutum cDNA, mRNA sequence.
ACCESSION AF373211
VERSION AF373211.1 GI:20428752
KEYWORDS EST.
SOURCE upland cotton.
ORGANISM Gossypium hirsutum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 571)
AUTHORS Li,J., Zhao,G. and Liu,J.
TITLE Isolation and characterization of cDNA fragments induced by salicylic acid in cotton
JOURNAL Unpublished (2002)
COMMENT Contact: Liu Jinyuan
Biological Science and Biotechnology
Tsinghua University
Haidian, Beijing, 100084, China
Email: Liujy@mail.tsinghua.edu.cn.
FEATURES Location/Qualifiers
source 1..571
/organism="Gossypium hirsutum"
/db_xref="taxon:3635"
/clone_lib="Gossypium hirsutum cotyledon"
/tissue_type="cotyledon"

```
                     /note="induced by salicylic acid"
BASE COUNT      147 a    133 c    150 g    141 t
ORIGIN

  Query Match           29.1%;  Score 351;  DB 9;  Length 571;
  Best Local Similarity  100.0%;  Pred. No. 6e-92;
  Matches  351;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     351 ATGAACGTTTTTAATCCCGCGCAGTTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 292

Qy      61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     291 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 232

Qy     121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     231 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 172

Qy     181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 112

Qy     241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     111 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 52

Qy     301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCA 351
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      51 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCA 1


RESULT 2
CNS01026
LOCUS       CNS01026                656 bp    DNA     linear   GSS 14-JUN-2001
DEFINITION  Anopheles gambiae GSS T7 end of clone 25F18 of NotreDame1 library
            from strain PEST of Anopheles gambiae (African malaria mosquito),
            genomic survey sequence.
ACCESSION   AL152975
VERSION     AL152975.1  GI:7013894
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
REFERENCE   2  (bases 1 to 656)
  AUTHORS   Roth,C.W., Brey,P.T., Ke,Z., Collins,F.H. and Weissenbach,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-FEB-2000) BBMI, Institut Pasteur, 25, rue du Dr.
            Roux, Paris 75015, France
COMMENT     This clone is from an A. gambiae BAC library provided by F.H.
            Collins and sequenced by Genoscope in collaboration with the
            Laboratory of Biochem. and Biol. Molec. of Insects, Institut
            Pasteur.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="25F18"
                     /clone_lib="NotreDame1"
                     /note="end : T7"
BASE COUNT      124 a    169 c    218 g    134 t     11 others
ORIGIN

  Query Match           27.6%;  Score 333.4;  DB 17;  Length 656;
  Best Local Similarity  73.3%;  Pred. No. 9.5e-87;
  Matches  471;  Conservative    5;  Mismatches  160;  Indels    7;  Gaps    4;

Qy     567 CGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCA 626
           ||| |||||||||||||||||| || || ||||| |  |||||||||||||| || || ||
Db       6 CGGTATGGTGGTGATGGTTGACGGCGCCCAGGGCGTGGTGCATTTCCCCGCCGACGTGCA 65

Qy     627 GCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTAT 686
           |   |||||||| ||,|||||||| || ||||| || || || || || || ||| |||||
Db      66 GGCGCTGGATATCGACTTCTATGCCTTCTCAGGACATAAGCTCTACGGGCCAACCGGTAT 125

Qy     687 CGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGG 746
           || ||| |||| || ||  | ||||| |||| |||  | || ||||| || |||||
Db     126 TGGGGTGTTGTACGGCAAGCCGGAACTTCTGGCAAAAATGACCCCGTGGCTCGGGGGCGG 185

Qy     747 CAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACT 806
           ||| ||| |    ||||||| ||||||||| |||  || ||   || ||| |     ||
Db     186 CAAGATGATCGCTGAAGTGACTTTTGACGGTTTCAAAACGCAGGATGTGCCTTACCGCCT 245

Qy     807 GGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCT 866
           ||||| || |||| || || || || || || || ||  |||||||| ||||| |||||
Db     246 TGAAGCCGGTACGCCGAACGTGGCCGGCGTGATTGGCCTAAGCGCGGNGCTGGAGTGGCT 305

Qy     867 GGCAGATTACGATATCAACCAGGCCGAAAGCTGGAGCC--GTAGCTTAGCAACGCTGGCG 924
           || |||  | || |     ||||  ||||:|||||| |  | | | |  | |||||||||
Db     306 GGAAGAAAAAGACAGTGTWCAGGNTGAAARCTGGAGACGGGAGGNTAGGAAACGCTGGCA 365

Qy     925 GAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCATTCCGCTGCCAGGATTCCAGCCTG 984
           || || |   ||   || || || || || || || || ||||||  ||||||||| ||||||
Db     366 GAGGAAGAAATGAAAAAGCGCCCGGGTTTCCGNTCGTTCCGCGTTCAGGATTCAAGCCTG 425

Qy     985 CTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGATATGGTGACGCTGCTGGCGGAGTAC 1044
           || |||||||||||||  ||||| ||||| ||||| |||||:|||||||||| | |||
Db     426 CTCGCCTTTGATTTTGTGGGCGTGCATCACAGCGACATGGTGMCGCTGCTGGCCGGGTAT 485

Qy    1045 GGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGTA 1104
           || || |||||||| |||||||||| ||||| ||  |||| |||||  |   | |||||
Db     486 GGCATCGCCCTGCGCGCCGGGCAGCACTGCGCGCA--CGCTGCTGGCGGCGCTCGGCGTG 543

Qy    1105 ACCGGCACACTGCGCGCCTCTTTTGCGCCATATAATACAAAGAGTGATGTGGATGCGCTG 1164
           ||  |    :|||||||||| || ||| | ||||| ||| | :| ||||| || ||||||
Db     544 AC--GGCACSTGCGCGCCTCGTTCGCGTCCTATAACACACAAMGCGATGTTGACGCGCTG 601

Qy    1165 GTGAATGCCG-TTGACCGCGCGCTGGAATTATTGGTGGATTAA 1206
           ||    |||| |||||||||||||| ||| || |||||:||||
Db     602 GTTGCCGCCGTTTGACCGCGCGCTTGAAATACTGGTGGMTTAA 644


RESULT 3
AF095401/c
LOCUS       AF095401                932 bp    DNA     linear   GSS 29-AUG-2000
DEFINITION  AF095401 Salmonella typhimurium LT2, Lambda DASH II Salmonella
            typhimurium genomic clone 77-T7, DNA sequence.
ACCESSION   AF095401
VERSION     AF095401.1  GI:4322852
KEYWORDS    GSS.
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 932)
  AUTHORS   Wong,R.M.Y. and McClelland,M.
  TITLE     End Sequences of Salmonella typhimurium LT2 Lambda DASHII Clones,
            Li-Cor
  JOURNAL   Unpublished (1999)
COMMENT     Contact: McClelland M
            Molecular Biology
            Sidney Kimmel Cancer Center
            3099 Science Park Road, San Diego, CA 92121, USA
            Email: mcclelland@lifsci.sdsu.edu
            Class: shotgun
FEATURES             Location/Qualifiers
```

```
     source          1. .932
                     /organism="Salmonella typhimurium"
                     /strain="LT2"
                     /db_xref="taxon:602"
                     /clone="77-T7"
                     /clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
                     /note="Vector: Lambda DASH II; sequenced using Li-Cor
                     sequencer"
BASE COUNT      220 a    234 c    262 g    215 t      1 others
ORIGIN

  Query Match           26.5%;  Score 320;  DB 17;  Length 932;
  Best Local Similarity   76.5%;  Pred. No. 9.7e-83;
  Matches  406;  Conservative    0;  Mismatches  121;  Indels    4;  Gaps    1;

Qy      1 ATGAACGTTTTTAATCCCGCGCAGTTCGCGCCCAGTTTCCCGCACTACAGGATGCGGGC 60
          ||||||| | |||||||| |||||||||||||| |||||||||| |||   ||||||||
Db    531 ATGAACGCTCTTAATCCCACGCAGTTTCGCGCGCAGTTTCCCGCGCTAGCCGATGCGGGT 472

Qy     61 GTCTATCTCGACAGCGCCGCGACCGCGCTTAAACCTGAAGCCGTGGTTGAAGCCACCCAA 120
          || |||||||| ||||||| || ||  | || ||  | || ||  |||| ||||| ||
Db    471 GTTTATCTCGATAGCGCCGCCACGGCATTAAAGCCACAGGCAGTCATTGACGCCACGCAC 412

Qy    121 CAGTTTTACAGTCTGAGCGCCGGAAACGTCCATCGCAGCCAGTTTGCCGAAGCCCAACGC 180
          ||||||||  || |||||||||| ||||| ||||| ||||||||||  | || |    |
Db    411 CAGTTTTATTGTTTGAGCGCCGGTAACGTTCATCGTAGCCAGTTTGCGCAGGCGC----C 356

Qy    181 CTGACCGCGCGTTATGAAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGAT 240
           |||| ||||  |||||||| ||  |||  ||||  ||||   |  | || || || |||
Db    355 GTGACGGCGCAATATGAAGCGGCCAGAGCAAAAGCAGCGCGGCTGTTAAACGCGCCCGAT 296

Qy    241 GATAAAACTATCGTCTGGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGC 300
          || |||| |||||||||||| |||||||||| ||| | ||||||||||||| || ||
Db    295 GAAAAAAGTATCGTCTGGACACGCGGCACCACCGAAGCGATCAACATGGTGGCGCAGTGT 236

Qy    301 TATGCGCGTCCGCGTCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCAC 360
          || || ||||| ||||||| || |||||||| ||||| || ||||| || || || |||
Db    235 TACGCCCGTCCTCGTCTGCGCCCCGGCGATGAAATTATCGTTAGCGTCGCCGAGCATCAC 176

Qy    361 GCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTG 420
          |||||||| || |||||||||||||| || |||||||| || ||  | ||  | | | ||
Db    175 GCCAACCTTGTGCCCTGGCTGATGGTGGCGCAACAAACCGGCGCGCAGGTCATAACACTG 116

Qy    421 CCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGT 480
          ||||||||||  | ||| || || |||||  ||     |||| ||||| || ||  | ||
Db    115 CCGCTTAATGACCGGCGTCTTCCTGATGTTGAGCGTCTGCCGGAACTCATCACGTCGCGC 56

Qy    481 AGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCG 531
          || |||| ||||||| |||| || ||||||||||    ||||| ||||||
Db     55 AGCCGGACTCTGGCGCTGGGGCAAATGTCGAACGTCNACGGCGGCTGCCCG 5


RESULT 4
BQ518982
LOCUS       BQ518982                 755 bp    mRNA    linear   EST 10-JUN-2002
DEFINITION  EST626397 Generation of a set of potato cDNA clones for microarray
            analyses mixed potato tissues Solanum tuberosum cDNA clone STMJN24
            5' end, mRNA sequence.
ACCESSION   BQ518982
VERSION     BQ518982.1  GI:21377851
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 755)
  AUTHORS   Buell,C.R., Hart,A., Baker,B., Tanksley,S., Fry,W., Smart,C.,
            Restrepo,S., Griffiths,H., van der Hoeven,R., Tsai,J. and
            Karamycheva,S.A.
  TITLE     Generation of a set of potato cDNA clones for microarray analyses
  JOURNAL   Unpublished (2002)
COMMENT     Other_ESTs: EST626398
            Contact: Robin Buell
            The Institute for Genomic Research
            9712 Medical Center Dr, Rockville, MD 20850, USA
            Email: potato@tigr.org
            This clone is available through the Research Genetics, contact the
            Research Genetics for further information 1-800-711-6195 or
            cdna@resgen.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .755
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec or Binjte"
                     /db_xref="taxon:4113"
                     /clone="STMJN24"
                     /clone_lib="Generation of a set of potato cDNA clones for
                     microarray analyses mixed potato tissues"
                     /tissue_type="mixed tissues"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Combination of untreated and Phytophthora
                     infestans-treated libraries of stolons, leaves, leaflets,
                     axillary buds of stem explants, petioles, germinating eyes
                     , tubers, or roots."
BASE COUNT      190 a    141 c    183 g    241 t
ORIGIN

  Query Match           11.5%;  Score 138.4;  DB 14;  Length 755;
  Best Local Similarity   51.5%;  Pred. No. 2e-29;
  Matches  319;  Conservative    0;  Mismatches  301;  Indels    0;  Gaps    0;

Qy    269 CCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCGTCCGCGTCTGCAACCGGCG 328
          | |||||  | |||||| | |||||  | |  |  |  |    || |||| || |
Db     11 CTACTGAGGCTATCAACTTAGTGGCTTATTCTTGGGGACTCTCAAACTTGAAACCAGGAG 70

Qy    329 ATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCG 388
          ||||||||||   ||   | || ||||| |||        | || || |||| || || |
Db     71 ATGAGATTATAACCACAATAGCTGAACATCACAGTGCAATTGTTCCTTGGCAAATTGTTG 130

Qy    389 CCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATG 448
          | ||| | ||||| ||    ||  |    |     | |  | | |   | | || ||||
Db    131 CTCAAAAGACTGGTGCTGTCCTGAGGTTTGTAAATTTGACAGAGGATGAAGTTCCAGATG 190

Qy    449 TCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGT 508
          | |     ||  |||  |  |  |     | |      || |||  ||  |||  | |
Db    191 TAGGGCAATTAAGAGACTTTCTCTCAAAGAGCACAAAACTTTTGGTCATGCATCATGTCT 250

Qy    509 CGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCG 568
          ||||  |    |    |  |  | |||    ||  || | |  || |||   | |
Db    251 CGAATATGTTAGCTTCTGTTCTTCCTATGGATGAAGTCATAAGTTGGGCACATGATGTTG 310

Qy    569 GGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGC 628
          |       ||| | || |||| |  |||  | |     ||  |    |  |||||| ||
Db    311 GAGCAAAAGTGCTTGTAGATGCTTGTCAAAGTGTCCCACACATGGTGGTCGATGTTAAGA 370

Qy    629 AACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCG 688
          | || || | ||||||   || |  |||   |||  ||  ||| ||| || |||||  | |
Db    371 ATCTAGACACTGATTTTCTTGTTGGTTCCTCTCATAAGATGTGTGGGCCTACAGGCGTTG 430

Qy    689 GCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCA 748
          |  |  ||||||| ||    ||  | ||    |  ||| | || |  || || |||
Db    431 GATTCTTGTATGGAAAGAGCGAGATCCTTTCTGGCATGCCTCCTTTCCTTGGTGGTGGCG 490

Qy    749 AAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGG 808
          ||||| |    || |||  ||| ||     | |||   |        | || | || | | |
Db    491 AAATGATAGCTGATGTGTATTTGGATCATTCCACTTATGCTGAACCTCCTTCGAGATTTG 550

Qy    809 AAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCTGG 868
          | ||||| || ||     | |  |  |  || ||  |  | || ||  | || |  ||
Db    551 AGGCTGGGACTCCCGCAATTGGAGAAGCAATTGGGCTTGGAGCTGCAATTGATTATCTTT 610

Qy    869 CAGATTACGATATCAACCAG 888
          | ||    |  ||  | |||
Db    611 CTGAAATTGGCATGCAACAG 630
```

```
RESULT 5
BM407607
LOCUS       BM407607                725 bp    mRNA    linear   EST 22-JAN-2002
DEFINITION  EST581922 potato roots Solanum tuberosum cDNA clone cPRO31L21 5'
            end, mRNA sequence.
ACCESSION   BM407607
VERSION     BM407607.1  GI:18259225
KEYWORDS    EST.
SOURCE      potato.
  ORGANISM  Solanum tuberosum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
REFERENCE   1  (bases 1 to 725)
  AUTHORS   van der Hoeven,R., Sun,H., Karamycheva,S.A., Tsai,J., Van Aken,S.,
            Utterback,T., Chiemingo,A., Bougri,O., Buell,C.R., Ronning,C.,
            Tanksley,S. and Baker,B.
  TITLE     Generation of ESTs from potato roots
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Research Genetics, Libraries Division
            Tel: 1-800-711-6195
            Email: cdna@resgen.com
            For clone info: please contact Research Genetics, Libraries
            Division tel 1-800-711-6195, email cdna@resgen.com
            Seq primer: T3.
FEATURES             Location/Qualifiers
     source          1. .725
                     /organism="Solanum tuberosum"
                     /cultivar="Kennebec"
                     /db_xref="taxon:4113"
                     /clone="cPRO31L21"
                     /clone_lib="potato roots"
                     /tissue_type="roots"
                     /dev_stage="in vitro grown stem cuttings"
                     /lab_host="SOLR"
                     /note="Vector: pBluescript SK(-); Site_1: EcoRI; Site_2:
                     XhoI; supplier: Cornell University, Tanksley lab;
                     sequencing; The Institute for Genomic Research. Roots were
                     isolated from in vitro grown stem cuttings on CM medium.
                     Roots were isolated two weeks after placing the stem
                     cuttings from in vitro grown plants on medium."
BASE COUNT      184 a    135 c    177 g    229 t
ORIGIN

  Query Match             11.3%;  Score 136.8;  DB 13;  Length 725;
  Best Local Similarity   51.7%;  Pred. No. 5.7e-29;
  Matches  312;  Conservative    0;  Mismatches  292;  Indels    0;  Gaps    0;

Qy     269 CCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCGTCCGCGTCTGCAACCGGGCG 328
           | |||||  | |||||| | |||||  | |  |  |  |  |    || |||| || |
Db      11 CTACTGAGGCTATCAACTTAGTGGCTTATTCTTGGGGACTCTCAAACTTGAAACCAGGAG 70

Qy     329 ATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCG 388
           ||||||||||   ||   | || |||||  |||      | || || ||||  || || |
Db      71 ATGAGATTATAACCACAATAGCTGAACATCACAGTGCAATTGTTCCTTGGCAAATTGTTG 130

Qy     389 CCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATG 448
           | ||| | ||||| ||     ||  |    |     | |  | | |   | | || ||||
Db     131 CTCAAAAGACTGGTGCTGTCCTGAGGTTTGTAAATTTGACAGAGGATGAAGTTCCAGATG 190

Qy     449 TCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGT 508
           | |     ||   |||  |  |  |  | |      || |||   ||  |||  | |
Db     191 TAGGGCAATTAAGAGACTTTCTCTCAAAGAGCACAAAACTTTTGGTCATGCATCATGTCT 250

Qy     509 CGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCG 568
           ||||  |    |   |   |   | |||    ||  || | |  || |||   |  |
Db     251 CGAATATGTTAGCTTCTGTTCTTCCTATGGATGAAGTCATAAGTTGGGCACATGATGTTG 310

Qy     569 GGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGC 628
           |      ||| | || |||| |  |||  | |     ||  |    |  ||||||  ||
Db     311 GAGCAAAAGTGCTTGTAGATGCTTGTCAAAGTGTCCCACACATGGTGGTCGATGTTAAGA 370

Qy     629 AACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCG 688
           | || || | ||||||   || |  |||   ||| ||  ||| ||| || |||||  | |
Db     371 ATCTAGACACTGATTTTCTTGTTGGTTCCTCTCATAAGATGTGTGGGCCTACAGGCGTTG 430

Qy     689 GCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCA 748
           |  |  |||||||  ||    ||  | ||    |  ||| | || |  || || || |||
Db     431 GATTCTTGTATGGAAAGAGCGAGATCCTTTCTGGCATGCCTCCTTTCCTTGGTGGTGGCG 490

Qy     749 AAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGG 808
           |||||  |    || |||  |||  ||    | |||   |      | || | || | | |
Db     491 AAATGATAGCTGATGTGTATTTGGATCATTCCACTTATGCTGAACCTCCTTCGAGATTTG 550

Qy     809 AAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGCTGG 868
           | |||||  || ||    | |  |  |  |  || ||  |  | || ||  | || |  ||
Db     551 AGGCTGGGACTCCCGCAATTGGAGAAGCAATTGGGCTTGGAGCTGCAATTGATTATCTTT 610

Qy     869 CAGA 872
           | ||
Db     611 CTGA 614


RESULT 6
BQ578320
LOCUS       BQ578320                645 bp    mRNA    linear   EST 19-JUN-2002
DEFINITION  WHE0301_E11_I21ZS Wheat unstressed seedling shoot cDNA library
            Triticum aestivum cDNA clone WHE0301_E11_I21, mRNA sequence.
ACCESSION   BQ578320
VERSION     BQ578320.1  GI:21481637
KEYWORDS    EST.
SOURCE      bread wheat.
  ORGANISM  Triticum aestivum
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
            ; Triticeae; Triticum.
REFERENCE   1  (bases 1 to 645)
  AUTHORS   Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
            ,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
            Seaton,C.L. and Tong,J.C.
  TITLE     The structure and function of the expressed portion of the wheat
            genomes - Etiolated shoot cDNA library
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Olin Anderson
            US Department of Agriculture, Agriculture Research Service, Pacific
            West Area, Western Regional Research Center
            800 Buchanan Street, Albany, CA 94710, USA
            Tel: 5105595773
            Fax: 5105595818
            Email: oandersn@pw.usda.gov
            Sequences have been trimmed to remove vector sequence and low
            quality sequence with phred score less than 20
            Seq primer: SK primer.
FEATURES             Location/Qualifiers
     source          1. .645
                     /organism="Triticum aestivum"
                     /cultivar="Chinese Spring"
                     /db_xref="taxon:4565"
                     /clone="WHE0301_E11_I21"
                     /clone_lib="Wheat unstressed seedling shoot cDNA library"
                     /tissue_type="Etiolated shoot"
                     /dev_stage="Five day old seedling"
                     /lab_host="E. coli SOLR"
                     /note="Vector: Lambda Uni-ZAP XR, excised phagemid;
                     Site_1: EcoRI; Site_2: XhoI; Seeds were surface-sterilized
                     , germinated and grown aseptically in the dark at room
                     temperature on filter paper with water, nystatin and
                     cefotaxime in covered crystallization dishes. Shoots were
                     harvested.  The tissue, total RNA, and poly(A) RNA were
                     prepared, a cDNA library  was made, and the cDNA clones
                     were in vivo excised to give pBluescript  phagemids in the
                     TJ Close lab (Choi, Close, Fenton) at the University  of
                     California, Riverside. Plasmid DNA preparations and DNA
                     sequencing  were performed in the OD Anderson lab (all
```

other authors)."
BASE COUNT 175 a 114 c 157 g 199 t
ORIGIN

Query Match 11.2%; Score 135.4; DB 14; Length 645;
Best Local Similarity 51.0%; Pred. No. 1.4e-28;
Matches 319; Conservative 0; Mismatches 306; Indels 0; Gaps 0;

```
Qy    197 AAGCTGCACGAGAGAAAGTGGCGCAATTACTGAATGCACCGGATGATAAAACTATCGTCT 256
          | |  ||| ||  |||  | ||  | ||  | |||||  | |||  || |   ||  | |
Db      6 ACGAGGCAAGAAGGAATCTTGCAAATTTCATCAATGCGGCTGATAGTAGAGAGATTATTT 65

Qy    257 GGACGCGCGGCACCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCGTCCGCGTC 316
            || ||     | || ||| | |||||  | || ||  | |  |  |  |   |
Db     66 TTACTCGTAATGCTACAGAAGCTATCAATTTAGTAGCTTATTCATGGGGCCTATCTAACT 125

Qy    317 TGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCT 376
          |  |||  || |||||||||||| | |  || || || || ||        | || || |
Db    126 TAAAACAAGGAGATGAGATTATTCTTACAGTTGCGGAGCATCATAGTGCGATTGTTCCTT 185

Qy    377 GGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGC 436
          |||   | ||  ||||| |||||| ||      |  | |  |  ||| | |    |
Db    186 GGCAATTTGTATCCCAAAAAACTGGTGCTGTCCTAAAGTATGTTGGGCTGACTAAAGAAG 245

Qy    437 GACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGT 496
            | || ||  | ||   |||   ||  ||  |  |   |   |    |||| | |
Db    246 AGGTTCCAGACATTGAGCAGTTAAAAGGGCTTCTGTCAAGCAAGACAAAGATTGTTGTTG 305

Qy    497 TGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTG 556
          |  |||  | || |||||    ||        |    | | | |   |   | | | ||
Db    306 TCCATCATGTTTCAAACGTCCTGGGTTCAATGCTTCCTATTGAGGAGGTCGTATCATGTG 365

Qy    557 CTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCG 616
          |  |   ||  ||      ||  | || |||| |  |||  | |     ||| | || |
Db    366 CAAACAAAGTTGGAGCTAAAGTTCTTGTAGATGCTTGTCAAAGTGTTCCTCATATGCCAG 425

Qy    617 CGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCC 676
            |||||||||  | ||  | |   |||||   || |   ||    || ||  ||| |||||
Db    426 TTGATGTTCAGAAGCTTGGTGCAGATTTTCTTGTTGCATCCTCGCATAAGATGTGTGGCC 485

Qy    677 CGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGC 736
          | |||||  ||||  |  || |||| ||||  ||  |  ||    | |||  ||| |
Db    486 CTACAGGCGTCGGATTCTTGCATGGAAAATTTGAGATCTTGTCATCTATGGAGCCTTTCT 545

Qy    737 TGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGC 796
          | || || ||  ||||| ||   || |||       |||   | ||   |       | |
Db    546 TAGGTGGTGGTGAAATGATTGCAGATGTGTTCGAAGACAAATCTACATATGCTGAGCCTC 605

Qy    797 CGTGGAAACTGGAAGCTGGAACGCC 821
          | |  | | | || ||||||||| ||
Db    606 CTTCTAGATTTGAGGCTGGAACTCC 630
```

RESULT 7
AV550366
LOCUS AV550366 552 bp mRNA linear EST 06-SEP-2000
DEFINITION AV550366 Arabidopsis thaliana roots Columbia Arabidopsis thaliana cDNA clone RZ111d02R 5', mRNA sequence.
ACCESSION AV550366
VERSION AV550366.1 GI:8721779
KEYWORDS EST.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 552)
AUTHORS Asamizu,E., Nakamura,Y., Sato,S. and Tabata,S.
TITLE A large scale analysis of cDNA in Arabidopsis thaliana: Generation of 12,028 non-redundant expressed sequence tags from normalized and size-selected cDNA libraries
JOURNAL DNA Res. 7, 175-180 (2000)
MEDLINE 20363093

COMMENT Contact: Erika Asamizu
The First Laboratory for Plant Gene Research
Kazusa DNA Research Institute
Yana 1532-3, Kisarazu, Chiba 292-0812, Japan
Email: asamizu@kazusa.or.jp, URL:http://www.kazusa.or.jp/en/plant/.
FEATURES Location/Qualifiers
source 1..552
/organism="Arabidopsis thaliana"
/strain="Columbia"
/db_xref="taxon:3702"
/clone="RZ111d02R"
/clone_lib="Arabidopsis thaliana roots Columbia"
/tissue_type="roots"
/note="Vector: pBluescriptII SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 145 a 106 c 123 g 178 t
ORIGIN

Query Match 10.6%; Score 127.8; DB 10; Length 552;
Best Local Similarity 53.3%; Pred. No. 2.2e-26;
Matches 270; Conservative 0; Mismatches 237; Indels 0; Gaps 0;

```
Qy    315 TCTGCAACCGGGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCC 374
          |||  |||| || || ||| ||||  | |  || || |||||  ||       | | || ||
Db     23 TCTTAAACCAGGAGACGAGGTTATACTTACAGTTGCTGAACATCATAGTTGCATTGTTCC 82

Qy    375 CTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCA 434
          ||||  || ||  | ||| ||||||||||      |   |   || |  | |||| | |
Db     83 TTGGCAAATTGTATCTCAAAAAACTGGAGCAGTTCTAAAGTTTGTGACTTTGAATGAGGA 142

Qy    435 GCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGC 494
            | | || ||  |  |  ||||  |||| ||||| ||||   |    | |  | ||
Db    143 TGAAGTCCCTGACATAAACAAGTTGAGAGAATTGATTTCTCCAAAGACAAAGCTCGTCGC 202

Qy    495 GTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTT 554
            |   |||  | || || ||| ||    ||  |    | ||           |    |
Db    203 TGTTCATCATGTTTCTAATGTTCTTGCATCTTCTCTTCCTATTGAAGAGATTGTGGTATG 262

Qy    555 TGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCC 614
          || ||     |  ||   |   ||  | || |||| |  |||  | |     ||  |
Db    263 GGCACACGATGTTGGAGCGAAAGTTCTAGTGGATGCTTGTCAAAGTGTTCCTCACATGGT 322

Qy    615 CGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGG 674
            |  ||||| ||| | ||  ||  ||||||    | |   ||  |  ||||||  ||| |||
Db    323 GGTTGATGTCCAGAAGCTTAATGCTGATTTTCTAGTTGCGTCTTCTCACAAGATGTGTGG 382

Qy    675 CCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTG 734
          || ||||| || ||  |  | ||||||||     || || ||  |  | ||| | || |
Db    383 ACCTACAGGCATTGGGTTCTTATATGGTAAGAGTGATCTTCTACATTCCATGCCTCCATT 442

Qy    735 GCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGC 794
          | || |||||  |||||  |     || ||   | |||||     |  ||    |     |
Db    443 CTTAGGTGGCGGAGAAATGATCTCAGATGTATTTCTTGACCATTCAACTTATGCGGAACC 502

Qy    795 GCCGTGGAAACTGGAAGCTGGAACGCC 821
          || |  | | | ||||||||||| ||
Db    503 TCCATCCAGATTTGAAGCTGGAACACC 529
```

RESULT 8
AV914366
LOCUS AV914366 656 bp mRNA linear EST 18-JAN-2002
DEFINITION AV914366 K. Sato unpublished cDNA library, cv. Haruna Nijo germination shoots Hordeum vulgare subsp. vulgare cDNA clone bags5n20 5', mRNA sequence.
ACCESSION AV914366
VERSION AV914366.1 GI:18210143
KEYWORDS EST.
SOURCE Hordeum vulgare subsp. vulgare.
ORGANISM Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae ; Triticeae; Hordeum.

```
REFERENCE   1  (bases 1 to 656)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE     Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT     Contact: Tadasu Shin-i
            Center For Genetic Resource Information
            National Institute of Genetics
            1111 Yata, Mishima, Shizuoka 411-8540, Japan
            Tel: 81-559-81-6856
            Fax: 81-559-81-6855
            Email: tshini@genes.nig.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .656
                     /organism="Hordeum vulgare subsp. vulgare"
                     /cultivar="Haruna Nijo"
                     /db_xref="taxon:112509"
                     /clone="bags5n20"
                     /clone_lib="K. Sato unpublished cDNA library, cv. Haruna
                     Nijo germination shoots"
                     /tissue_type="shoots"
                     /dev_stage="germination"
BASE COUNT      163 a    118 c    170 g    205 t
ORIGIN

  Query Match             10.5%;  Score 126.6;  DB 10;  Length 656;
  Best Local Similarity   51.7%;  Pred. No. 5.4e-26;
  Matches  288;  Conservative    0;  Mismatches  269;  Indels    0;  Gaps    0;

Qy    325 GGCGATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGATG 384
          || |||||||||||| | |  || || || || ||        | || || ||||   |
Db      3 GGAGATGAGATTATTCTTACAGTTGCGGAGCATCATAGTGCAATTGTTCCTTGGCAATTT 62

Qy    385 GTCGCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCG 444
          ||  ||||| |||||||| ||     |   | |  |   ||| | |     |     | ||
Db     63 GTATCCCAAAAAACTGGTGCTGTCCTAAAGTATGTTGGGCTGACTAAAGAAGAGGTTCCA 122

Qy    445 GATGTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAG 504
          ||  | ||   |||   ||  ||  |  |   |   |    |||| | |   |   |||
Db    123 GACATTGAGCAGTTAAAAGGGCTTCTGTCAAGCAAGACCAAGATTGTTGTTGTCCATCAT 182

Qy    505 ATGTCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCA 564
          | ||||||||    ||       |    | | | |   |   |  | | |||  |   |
Db    183 GTTTCGAACGTCCTGGGTTCAATGCTTCCTATTGAGGAGGTTGTATCATGTGCAAACAAA 242

Qy    565 GCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTT 624
          |  ||        ||  | || |||| |  |||  | |     ||| | || |  ||||||
Db    243 GTTGGAGCTAAAGTTCTTGTAGATGCTTGTCAAAGTGTTCCTCATATGCCAGTTGATGTT 302

Qy    625 CAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGT 684
          ||| | || | |   |||||   || |   ||    || ||  ||| |||||| |||||
Db    303 CAGAAGCTTGGTGCAGATTTTCTTGTTGCATCCTCGCATAAGATGTGTGGCCCTACAGGC 362

Qy    685 ATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGC 744
          ||||  |  || |||| ||||  ||  |  ||    | |||  ||| |   | || ||
Db    363 GTCGGATTCTTGCATGGAAAATTTGAGATCTTGTCATCTATGGAGCCTTTCTTAGGTGGT 422

Qy    745 GGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAA 804
          ||  ||||| ||   || |||      |||   |  ||   |       |||| |  | |
Db    423 GGTGAAATGATTGCAGATGTGTTAGAAGACAAATCTACATATGCTGAGCCGCCTTCTAGA 482

Qy    805 CTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGG 864
          | || |||||||| ||      | |  |  |  |||||||  | |  ||| | || |
Db    483 TTTGAGGCTGGAACTCCTGCAATTGGAGAAGCTATAGGATTGGGAGTTGCGATAGATTAT 542

Qy    865 CTGGCAGATTACGATAT 881
          ||| || | |  |  ||
Db    543 CTGTCACACTTTGGCAT 559


RESULT 9
BH242226
LOCUS       BH242226               713 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  ATZFC20TF ATZF Arabidopsis thaliana genomic clone ATZFC20, DNA
            sequence.
ACCESSION   BH242226
VERSION     BH242226.1  GI:16917246
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 713)
  AUTHORS   Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T20G10
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: ATZFC20TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1370. Caution: the DNA in this BAC may be from
            some non-Arabidopsis source
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .713
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="ATZFC20"
                     /clone_lib="ATZF"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      145 a    212 c    216 g    140 t
ORIGIN

  Query Match             10.2%;  Score 122.6;  DB 17;  Length 713;
  Best Local Similarity   50.5%;  Pred. No. 8.4e-25;
  Matches  327;  Conservative    0;  Mismatches  314;  Indels    6;  Gaps    1;

Qy    556 GCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCC 615
          || ||| ||  ||| ||    ||  |||| |||||    ||| | || |||||| |
Db      2 GCACATGCACGCGGCATTCCTGTTCTGGTCGATGGCAGCCAGAGCGCGGTGCATCTGAAT 61

Qy    616 GCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGC 675
          |||||  ||| |  | ||    || ||||  | |||| |  || |||||  | || ||
Db     62 ATCGATGTGCAGGATATCGACTGCGACTTCTTCGTTTTTACCGGGCACAAGATCTACGGA 121

Qy    676 CCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGG 735
          ||||| || || |||||||| ||||  || |  || || || | | ||||  |||||
Db    122 CCGACCGGCATTGGCGTGCTCTATGCCAAGTATGATCTTCTCGGCGGGATGCGGCCCTAT 181

Qy    736 CTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCG 795
          |||||||||  | ||| | | ||| ||||      ||  |  | ||          | ||
Db    182 AACGGCGGCGGCGAGATGATCCGCGAGGTGACCAAGGATTGGGTGACCTACGGTGATCCG 241

Qy    796 CCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCG 855
          ||   ||  | ||||| || || ||    | |  |  |  || ||  |  | || |||
Db    242 CCTCACAAGTTCGAAGCCGGCACCCCTGCAATTGTCGAGGCGATCGGCCTTGGTGCTGCG 301

Qy    856 CTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGGAGCCGTAGCTTAGCA 915
          | || |  |          ||  |  | |   ||   |      |   | |
Db    302 ATCGACTACGTCAATTCGATCGGCAAGGAGCGCATCGCCGCCCACGAGCACGACCTGTTG 361

Qy    916 ACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCATTCCGCTGCCAGGAT 975
          ||    ||| | |||   |||.  |           | ||   | || ||       |
Db    362 ACCTATGCGCACGATCGCCTGCGCGAGATCAATTCGCTGCGCATCATCGGCACGGCACAC 421

Qy    976 TCCAGCCTGCTGG------CCTTTGATTTTGCTGGCGTTCATCATAGCGATATGGTGACG 1029
          |    || ||      | || ||  |   |||| ||||   ||| | | |||
Db    422 GAAAAGGGGCCGGTGATCTCGTTCGAACTGAAGGGCGCGCATCCGCACGACGTCGCGACC 481
```

```
Qy   1030 CTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAGCCGCTACTG 1089
          |  | |    |   || || ||| | || |||||    |||||||    ||||||  ||
Db    482 GTCATCGACCGGGCGGGGATCGCCGTCCGCGCCGGCACTCATTGCGTGATGCCGCTTTTG 541

Qy   1090 GCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAATACAAAGAGT 1149
          | |   ||    || || | |||    || || || || |     ||||||||
Db    542 GAACGGTTCAATGTCACAGCCACGTGCCGGGCGTCATTCGGCATGTATAATACGCGTGCG 601

Qy   1150 GATGTGGATGCGCTGGTGAATGCCGTTGACCGCGCGCTGGAATTATT 1196
          || || ||    || | | | ||  |      |||| ||| || ||
Db    602 GAAGTCGACCAACTAGCGGAAGCGCTGATTAAGGCGCGGGATTTGTT 648


RESULT 10
BH242467
LOCUS       BH242467                 635 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  ATZFC60TR ATZF Arabidopsis thaliana genomic clone ATZFC60, DNA
            sequence.
ACCESSION   BH242467
VERSION     BH242467.1  GI:16917686
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 635)
  AUTHORS   Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T20G10
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: ATZFC60TF
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1370. Caution: the DNA in this BAC may be from
            some non-Arabidopsis source
            Seq primer: TR
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .635
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="ATZFC60"
                     /clone_lib="ATZF"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      127 a    187 c    196 g    125 t
ORIGIN

  Query Match             9.6%;  Score 116.2;  DB 17;  Length 635;
  Best Local Similarity   50.4%;  Pred. No. 6.1e-23;
  Matches  314;  Conservative    0;  Mismatches  303;  Indels    6;  Gaps    1;

Qy    556 GCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCC 615
          || ||| ||  ||| ||    ||  |||| |||||    ||| | || ||||||
Db      2 GCACATGCACGCGGCATTCCTGTTCTGGTCGATGGCAGCCAGAGCGCGGTGCATCTGAAT 61

Qy    616 GCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGC 675
          ||||| ||| |  | ||    || ||||  | |||| | || |||||  | || ||
Db     62 ATCGATGTGCAGGATATCGACTGCGACTTCTTCGTTTTTACCGGGCACAAGATCTACGGA 121

Qy    676 CCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGG 735
          ||||| || || |||||||| ||||  || |  || || || |  | ||||  |||||
Db    122 CCGACCGGCATTGGCGTGCTCTATGCCAAGTATGATCTTCTCGCCGGGATGCGGCCCTAT 181

Qy    736 CTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCG 795
          |||||||||  | ||| | | ||| ||||     ||  |  | ||          | ||
Db    182 AACGGCGGCGGCGAGATGATCCGCGAGGTGACCAAGGATTGGGTGACCTACGGTGATCCG 241

Qy    796 CCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCG 855
          ||    ||  | ||||| || || ||      | |  |  | ||| ||  |  | || |||
Db    242 CCTCACAAGTTCGAAGCCGGCACCCCTGCAATTGTCGAGGCGATCGGCCTTGGTGCTGCG 301

Qy    856 CTGGAATGGCTGGCAGATTACGATATCAACCAGGCCGAAAGCTGGAGCCGTAGCTTAGCA 915
          | || |   |          ||  |   | |    ||    |       |    | |
Db    302 ATCGACTACGTCAATTCGATCGGCAAGGAGCGCATCGCCGCCCACGAGCACGACCTGTTG 361

Qy    916 ACGCTGGCGGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCATTCCGC------TGC 969
          ||    ||| | |||   |||    |           | ||     || ||         |
Db    362 ACCTATGCGCACGATCGCCTGCGCGAGATCAATTCGCTGCGCATCATCGGCACGGCACAC 421

Qy    970 CAGGATTCCAGCCTGCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGATATGGTGACG 1029
          |  |         || |  | || ||  |    ||||  |||    |||  | | |||
Db    422 GAAAAGGGGCCGGTGATCTCGTTCGAACTGAAGGGCGCGCATCCGCACGACGTCGCGACC 481

Qy   1030 CTGCTGGCGGAGTACGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAGCCGCTACTG 1089
          |  | |    |   || || ||| | || |||||    |||||||    |||  |  ||
Db    482 GTCATCGACCGGGCGGGGATCGCCGTCCGCGCCGGCACTCATTGCGTGATGCCCGTTTTG 541

Qy   1090 GCAGAATTAGGCGTAACCGGCACACTGCGCGCCTCTTTTGCGCCATATAATACAAAGAGT 1149
          | |   ||    || || |||||    || || || || |    ||||||||
Db    542 GAACGGTTCAATGTCACAGGCACGTGCCGGGCGTCATTCGGCATGTATAATACGCGTGCG 601

Qy   1150 GATGTGGATGCGCTGGTGAATGC 1172
          || || ||    || | | | ||
Db    602 GAAGTCGACCAACTAGCGGAAGC 624


RESULT 11
BH374662/c
LOCUS       BH374662                 707 bp    DNA     linear   GSS 10-DEC-2001
DEFINITION  AG-ND-179H2.TR ND-TAM Anopheles gambiae genomic clone AG-ND-179H2,
            DNA sequence.
ACCESSION   BH374662
VERSION     BH374662.1  GI:17320804
KEYWORDS    GSS.
SOURCE      African malaria mosquito.
  ORGANISM  Anopheles gambiae
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
            Anopheles.
REFERENCE   1  (bases 1 to 707)
  AUTHORS   Shetty,J., Malek,J., Koo,H., Collins,F., Gardner,M. and Loftus,B.J.
  TITLE     Direct Submission of BAC-end sequences from Anopheles gambiae
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: AG-ND-179H2.TF
            Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            This clone is from an A. gambiae BAC library (ND-TAM) provided by
            F.H. Collins and sequenced by The Institute for Genomic Research
            (TIGR). The BAC library was generated from A. gambiae PEST strain
            DNA. All DNA was extracted from newly hatched first instar larvae
            to minimize the inclusion of DNA from microorganisms that inhabit
            the gut. The DNA is derived from mixed sexes of larvae. The BAC
            library was constructed at Texas A&M Universty BAC Center
            University, College Station, Texas 77843-2123, USA using a HindIII
            partial digest.
            Seq primer: M13 Rev
            Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1. .707
                     /organism="Anopheles gambiae"
                     /strain="PEST"
                     /db_xref="taxon:7165"
                     /clone="AG-ND-179H2"
                     /clone_lib="ND-TAM"
```

```
                     /note="Vector: pECBAC1; Site_1: HindIII"
BASE COUNT      200 a    158 c    122 g    227 t
ORIGIN

  Query Match             9.6%;  Score 115.2;  DB 17;  Length 707;
  Best Local Similarity   52.0%;  Pred. No. 1.3e-22;
  Matches  283;  Conservative    0;  Mismatches  258;  Indels    3;  Gaps    1;

Qy     328 GATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTC 387
           || ||||| ||| |          || || ||| | ||  | || || |||| |||  |
Db     606 GACGAGATCATTATTTCTTACCTGGAGCATCACTCTAATATTGTTCCTTGGCAGATATTG 547

Qy     388 GCCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGAT 447
             ||   ||| ||||| ||| |       | | ||| | ||||   |  |  |   | |
Db     546 TGTCAGAGAACGGGAGCAAAAGTTCGTGTAATTCCGATGAATGAAGAAGGTATTTTGCAA 487

Qy     448 GTCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATG 507
           | ||| |  ||   |||  | || ||      |     || |  |  |   |||| |
Db     486 ATTGATGTTCTGGATGAATGGCTTTCTGAGAAAACAAAACTTGTTTCTGTAAATCAGGTT 427

Qy     508 TCGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCC 567
           || || |     ||   |          | |    |   || |      |       ||
Db     426 TCCAATGCCCTAGGTATTGTTAATCCGATTGATGAAATCATCAGAAAAGTAAGAGCAAAA 367

Qy     568 GGGATGG---TGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTT 624
             |  |   | ||  |  | || ||||| ||||  |||   ||||||      |||||
Db     366 TCTAATGCTTTTGTATTCATAGACGGTGCGCAGGCTGCACCACATTTCGAAATAGATGTA 307

Qy     625 CAGCAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGT 684
           |||  | |||||  |||||||  ||||| |||||||| ||| |||| || || |||||
Db     306 CAGGCAATGGATTGCGATTTCTTCGCTTTCTCAGGTCATAAAATGTACGGACCAACAGGA 247

Qy     685 ATCGGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGC 744
           | |||  |  | |||||||  |||   |  | |||  | |    || |  |  || ||
Db     246 ACCGGAATTTTATATGGTAAGGCAGCCGTTTTAGAGCAGCTAAATCCATTCCACGGAGGT 187

Qy     745 GGCAAAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAA 804
           ||  ||||| || || |    |  |||||       ||   |  |      || |  | |
Db     186 GGAGAAATGATTGACCACTGTACATTTGAGAAAACAACCTATGCAGGACTTCCTTTCAGA 127

Qy     805 CTGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGG 864
           | ||||| ||||| ||||||  | || |||   || | |||  |  | ||| | || |
Db     126 TTTGAAGCCGGAACACCAAATATTGCAGGTAATATTGCATTGGGAACAGCGGTAGATTTT 67

Qy     865 CTGG 868
             |||
Db      66 ATGG 63


RESULT 12
BE511577
LOCUS       BE511577                 591 bp    mRNA    linear   EST 07-AUG-2000
DEFINITION  946062A01.y1 946 - tassel primordium prepared by Schmidt lab Zea
            mays cDNA, mRNA sequence.
ACCESSION   BE511577
VERSION     BE511577.1  GI:9732825
KEYWORDS    EST.
SOURCE      Zea mays.
  ORGANISM  Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1  (bases 1 to 591)
  AUTHORS   Walbot,V.
  TITLE     Maize ESTs from various cDNA libraries sequenced at Stanford
            University
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Walbot V
            Department of Biological Sciences
            Stanford University
            855 California Ave, Palo Alto, CA 94304, USA
            Tel: 650 723 2227
            Fax: 650 725 8221
            Email: walbot@stanford.edu
            Plate: 946062  row: A  column: 01.
FEATURES             Location/Qualifiers
     source          1..591
                     /organism="Zea mays"
                     /cultivar="OH43"
                     /db_xref="taxon:4577"
                     /clone_lib="946 - tassel primordium prepared by Schmidt
                     lab"
                     /tissue_type="tassels"
                     /dev_stage="just after the transition from vegetative to
                     inflorescence development"
                     /lab_host="XLOLR"
                     /note="Organ: tassels; Vector: HybriZAP; Site_1: EcoRI;
                     Site_2: XhoI; George Chuck dissected immature tassels
                     between 1mm and 3mm.  Sharon  Stanfield prepared the cDNA
                     library in HybriZAP.  Sample insert size  range was 350 bp
                     to 3 Kb with a 1 Kb average."
BASE COUNT      159 a    110 c    143 g    179 t
ORIGIN

  Query Match             9.5%;  Score 114.8;  DB 10;  Length 591;
  Best Local Similarity   49.8%;  Pred. No. 1.5e-22;
  Matches  290;  Conservative    0;  Mismatches  292;  Indels    0;  Gaps    0;

Qy     269 CCACTGAATCCATCAACATGGTGGCACAATGCTATGCGCGTCCGCGTCTGCAACCGGGCG 328
           | |||||| | || ||  |||| ||  | |  |  |  |  |     || ||   || |
Db      10 CTACTGAAGCTATTAATCTGGTAGCCTATTCATGGGGACTATCTAACTTGAAAGAAGGAG 69

Qy     329 ATGAGATTATTGTCAGCGTGGCAGAACACCACGCCAACCTCGTCCCCTGGCTGATGGTCG 388
           |||| ||| || | |  || ||||| || ||          | || || ||||   ||||
Db      70 ATGAAATTGTTCTTACAGTTGCAGAGCATCATAGTGCTATTGTTCCATGGCAATTGGTTT 129

Qy     389 CCCAACAAACTGGAGCCAAAGTGGTGAAATTGCCGCTTAATGCGCAGCGACTGCCGGATG 448
           ||||  | ||||| ||||   |   |  |  | | |    |     | || ||
Db     130 CCCAGAAGACTGGTGCCACCTTAAAATATGTTGGATTGACTAAAGAAAATGTCCCAGACA 189

Qy     449 TCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGT 508
           | ||   |||   ||  |  |  |   |   |   |||  | |   |   ||| | |
Db     190 TTGAGCAGTTAAAAGGGTTTCTGTCGAACAAGACAAAGATAGTTGTTGTTCATCATGTCT 249

Qy     509 CGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCG 568
           | || |||   ||        |   | | | |       | || |  |  |  |  || |
Db     250 CAAATGTTCTAGGTTCCATGCTTCCTATCGAGGAGATTGTAACATGGTCAAACAGAGTTG 309

Qy     569 GGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGC 628
           |       ||  | || |||| |   ||  | |     ||| | || |  |||||||||
Db     310 GAGCTAAAGTCCTTGTAGATGCTTGCCAAAGTGTTCCCCATATGCCAGTTGATGTTCAGA 369

Qy     629 AACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCG 688
           |||| |     || ||  ||     ||  ||| ||  ||| |||||| |||||  |||
Db     370 AACTTGGCGCGGACTTTCTTGTCGCATCCTCTCATAAGATGTGTGGCCCTACAGGCGTCG 429

Qy     689 GCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCA 748
           |  |  || |||||||||  ||  |  ||  | | |||  ||| |   | || || ||
Db     430 GATTTTTGCATGGTAAATTTGAGATTTTGTCGTCTATGGAGCCTTTCTTAGGTGGTGGTG 489

Qy     749 AAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGG 808
           ||||| ||   || ||         |||   | |||   |     | || |  | | | |
Db     490 AAATGATTGCGGATGTATTCGAAGACAAATCCACATATGCTGAACCTCCCTCTAGATTTG 549

Qy     809 AAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCG 850
           | |||||||| ||     |||  |  |||||| |  | |
Db     550 AGGCTGGAACTCCTGCAATCGGAGAAGCTATAGGACTGGGAG 591


RESULT 13
BH242443
LOCUS       BH242443                 735 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  ATZFD83TF ATZF Arabidopsis thaliana genomic clone ATZFD83, DNA
            sequence.
ACCESSION   BH242443
VERSION     BH242443.1  GI:16917649
```

```
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 735)
  AUTHORS   Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T20G10
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: ATZFD83TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1370. Caution: the DNA in this BAC may be from
            some non-Arabidopsis source
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .735
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="ATZFD83"
                     /clone_lib="ATZF"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      157 a    214 c    219 g    144 t      1 others
ORIGIN

  Query Match             9.5%;  Score 114.6;  DB 17;  Length 735;
  Best Local Similarity   51.7%;  Pred. No. 1.9e-22;
  Matches  261;  Conservative    0;  Mismatches  244;  Indels    0;  Gaps    0;

Qy    358 CACGCCAACCTCGTCCCCTGGCTGATGGTCGCCCAACAAACTGGAGCCAAAGTGGTGAAA 417
Db     76 CACTCCAACATCGTGCCGTGGAATTTCCTGCGCGAACGTCAGGGCGCGGTGATCAAATGG 135

Qy    418 TTGCCGCTTAATGCGCAGCGACTGCCGGATGTCGATTTGTTGCCAGAACTGATTACTCCC 477
Db    136 GCGCCTGTCGACGACGAAGGCAATTTCCTGATCGACGAATTTGAGAAGCTGCTGACGCCA 195

Qy    478 CGTAGTCGGATTCTGGCGTTGGGTCAGATGTCGAACGTTACTGGCGGTTGCCCGGATCTG 537
Db    196 AAGACCAAGCTGGTCGCCATCACCCAGATGTCGAATGCACTCGGCACCATCGTACCCGTG 255

Qy    538 GCGCGAGCGATTACCTTTGCTCATTCAGCCGGGATGGTGGTGATGGTTGATGGTGCTCAG 597
Db    256 AAGGACGTGGTGAAGATCGCACATGCACGCGGCATTCCTGTTCTGGTCGATGGCAGCCAG 315

Qy    598 GGGGCAGTGCATTTCCCCGCGGATGTTCAGCAACTGGATATTGATTTCTATGCTTTTTCA 657
Db    316 AGCGCGGTGCATCTGAATATCGATGTGCAGGATATCGACTGCGACTTCTTCGTTTTTACC 375

Qy    658 GGTCACAAACTGTATGGCCCGACAGGTATCGGCGTGCTGTATGGTAAATCAGAACTGCTG 717
Db    376 GGGCACAAGATCTACGGACCGACCGGCATTGGCGTGCTCTATGCCAAGTATGATCTTCTC 435

Qy    718 GAGGCGATGTCGCCCTGGCTGGGCGGCGGCAAAATGGTTCACGAAGTGAGTTTTGACGGC 777
Db    436 GCCGGGATGCGGCCCTATAACGGCGGCGGCGAGATGATCCGCGAGGTGACCAAGGATTGG 495

Qy    778 TTCACGACTCAATCTGCGCCGTGGAAACTGGAAGCTGGAACGCCAAATGTCGCTGGTGTC 837
Db    496 GTGACCTACGGTGATCCGCCTCACAAGTTCGAAGCCGGCACCCCTGCAATTGTCGAGGCG 555

Qy    838 ATAGGATTAAGCGCGGCGCTGGAAT 862
Db    556 ATCGGCCTTGGTGCTGCGATCGACT 580
```

```
RESULT 14
AF035972/c
LOCUS       AF035972               950 bp    DNA     linear   GSS 29-AUG-2000
DEFINITION  AF035972 Salmonella typhimurium LT2, Lambda DASH II Salmonella
            typhimurium genomic clone B324-T7, DNA sequence.
ACCESSION   AF035972
VERSION     AF035972.1  GI:2689043
KEYWORDS    GSS.
SOURCE      Salmonella typhimurium.
  ORGANISM  Salmonella typhimurium
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Salmonella.
REFERENCE   1  (bases 1 to 950)
  AUTHORS   Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
  TITLE     Sample sequencing of a Salmonella typhimurium LT2 lambda library:
            comparison to the Escherichia coli K12 genome
  JOURNAL   FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
  MEDLINE   99243757
COMMENT     Contact: McClelland M
            Molecular Biology
            Sidney Kimmel Cancer Center
            3099 Science Park Road, San Diego, CA 92121, USA
            Email: mcclelland@lifsci.sdsu.edu
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .950
                     /organism="Salmonella typhimurium"
                     /strain="LT2"
                     /db_xref="taxon:602"
                     /clone="B324-T7"
                     /clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
                     /note="Vector: Lambda DASH II; sequenced using Li-Cor
                     sequencer"
BASE COUNT      203 a    290 c    244 g    203 t     10 others
ORIGIN

  Query Match             9.5%;  Score 114.6;  DB 17;  Length 950;
  Best Local Similarity   51.8%;  Pred. No. 2.1e-22;
  Matches  329;  Conservative    0;  Mismatches  301;  Indels    5;  Gaps    3;

Qy    568 GGGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAG 627
Db    765 GGGGCGAAAGTGCTGNTGGATGGCGCCCAGGCCGTGATGCACCATGCTGTTGACGTCCAG 706

Qy    628 CAACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATC 687
Db    705 GCGCTGGACTGCGATTTTTACGTTCTCTCCGGCCATAAACTNTACGGGCCGACCGGCATC 646

Qy    688 GGCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGC 747
Db    645 GGCATTCTGTATGTTAAAGAGGCGTTGCTGCAAGAAATGCCGCCGTGGGAAGGGGGCGGG 586

Qy    748 AAAATGGTTCACGAAGTGAG--TTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAAC 805
Db    585 TCGATGATCTCGACCGTCAGCCTGACGCAGGGAACGACATGGCGAAAGCGCCCTGGCGTT 526

Qy    806 TGGAAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAATGGC 865
Db    525 TTGAGGCGGGAACGCCGAATACTGGCGGCATCATCGGTCTCGGCGCGGCGATTGATTATG 466

Qy    866 TGGCAGATTACGATATCAACCAGGCCGAAAGCTGGAGCCGTAGCTTAGCAACGCTGG--C 923
Db    465 TGACGTCGCTGGGACTGGATAAGATTGGCGATTATGAGCAGATGCTGATCGCTATCGCGCT 406

Qy    924 GGAAGATGCGCTGGCGAAACGTCCCGGCTTTCGTTCATTCCGCTGCCAGGATTCCAGCCT 983
Db    405 GGAGCAACTGGCGCAGGTGCCTGATATCACGCTATATGGCCCGGCGCAGCGGTTGGGCGT 346

Qy    984 GCTGGCCTTTGATTTTGCTGGCGTTCATCATAGCGATATGGTGACGCTGCTGGCGGAGTA 1043
Db    345 CATCGCGTTTAATCTGGGTAAACACCACGCTTACGACGTCGGCAGCTTTCTTGATAATTA 286
```

```
Qy   1044 CGGTATTGCCCTGCGGGCCGGGCAGCATTGCGCTCAGCCGCTACTGGCAGAATTAGGCGT 1103
          |||||| ||  | ||  | ||||| || |||||   ||||||  ||||    |  |||||
Db    285 CGGTATCGCGGTACGAACAGGGCATCACTGCGCAATGCCGCTCATGGCCTGGTATGGCGT 226

Qy   1104 AACCGGCACACTGCGCGCCTCTTTTGCGCCATATAATACAAAGAGTGATGTGGATGCGCT 1163
            ||||||   |||  || ||  | |||   ||||| ||  |    || |||||    ||
Db    225 -GCCGGCAATGTGCCGGCTTCGCTGGCGATGTATAACACCCATGAAGAAGTGGACCGACT 167

Qy   1164 GGTGAATGCCGTTGACCGCGCGCTGGAATTATTGG 1198
          ||||   |   |    ||    |     |||||||
Db    166 GGTGGCAGGATTAACGCGTATCCACCGCTTATTGG 132


RESULT 15
BH242236
LOCUS       BH242236                669 bp    DNA     linear   GSS 13-NOV-2001
DEFINITION  ATZFD22TF ATZF Arabidopsis thaliana genomic clone ATZFD22, DNA
            sequence.
ACCESSION   BH242236
VERSION     BH242236.1  GI:16917264
KEYWORDS    GSS.
SOURCE      thale cress.
  ORGANISM  Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE   1  (bases 1 to 669)
  AUTHORS   Kaul,S., Town,C.D., Bowman,C.L., Van Aken,S.E., Utterback,T.V.,
            Feldblyum,T.V. and Fraser,C.M.
  TITLE     Survey sequencing of Arabidopsis thaliana BAC T20G10
  JOURNAL   Unpublished (2001)
COMMENT     Other_GSSs: ATZFD22TR
            Contact: Chris Town
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA.
            Tel: 301-838-3523
            Fax: 301-838-0208
            Email: cdtown@tigr.org
            From Wash. U contig 1370. Caution: the DNA in this BAC may be from
            some non-Arabidopsis source
            Seq primer: TF
            Class: sheared ends.
FEATURES             Location/Qualifiers
     source          1. .669
                     /organism="Arabidopsis thaliana"
                     /strain="Columbia"
                     /db_xref="taxon:3702"
                     /clone="ATZFD22"
                     /clone_lib="ATZF"
                     /note="Vector: pHOS2; Site_1: BstXI; 2-3 kb sheared BAC
                     DNA inserted into pHOS2 using BstXI linkers"
BASE COUNT      141 a    197 c    202 g    127 t      2 others
ORIGIN

  Query Match             9.3%;  Score 111.6;  DB 17;  Length 669;
  Best Local Similarity   54.3%;  Pred. No. 1.4e-21;
  Matches  225;  Conservative    0;  Mismatches  189;  Indels    0;  Gaps    0;

Qy    449 TCGATTTGTTGCCAGAACTGATTACTCCCCGTAGTCGGATTCTGGCGTTGGGTCAGATGT 508
          ||||    ||     | ||| | || ||    |    | |  | ||  |    |||||||
Db     53 TCGACGAATTTGAGAAGCTGCTGACGCCAAAGACCAAGCTGGTCGCCATCACCCAGATGT 112

Qy    509 CGAACGTTACTGGCGGTTGCCCGGATCTGGCGCGAGCGATTACCTTTGCTCATTCAGCCG 568
          |||| |     |||     |       ||  |   | | | |   | || ||| ||  ||
Db    113 CGAATGCACTCGGCACCATCGTACCCGTGAAGGACGTGGTGAAGATCGCACATGCACGCG 172

Qy    569 GGATGGTGGTGATGGTTGATGGTGCTCAGGGGGCAGTGCATTTCCCCGCGGATGTTCAGC 628
          | ||    ||  |||| |||||    ||| | || ||||||  |       ||||| |||
Db    173 GCATTCCTGTTCTGGTCGATGGCAGCCAGAGCGCGGTGCATCTGAATATCGATGTGCAGG 232

Qy    629 AACTGGATATTGATTTCTATGCTTTTTCAGGTCACAAACTGTATGGCCCGACAGGTATCG 688
          |  | ||    || ||||  | |||| | || |||||  | || || ||||| || || |

Db    233 ATATCGACTGCGACTTCTTCGTTTTTACCGGGCACAAGATCTACGGACCGACCGGCATTG 292

Qy    689 GCGTGCTGTATGGTAAATCAGAACTGCTGGAGGCGATGTCGCCCTGGCTGGGCGGCGGCA 748
          ||||||| ||||  || |  || || || |  | ||||  |||||     |||||||||
Db    293 GCGTGCTCTATGCCAAGTATGATCTTCTCGCCGGGATGCGGCCCTATAACGGCGGCGGCG 352

Qy    749 AAATGGTTCACGAAGTGAGTTTTGACGGCTTCACGACTCAATCTGCGCCGTGGAAACTGG 808
          | ||| | | ||| ||||     ||  |  | ||         | ||||    ||  | |
Db    353 AGATGATCCGCGAGGTGACCAAGGATTGGGTGACCTACGGTGATCCGCCTCACAAGTTCG 412

Qy    809 AAGCTGGAACGCCAAATGTCGCTGGTGTCATAGGATTAAGCGCGGCGCTGGAAT 862
          |||| || || ||     | |  |  |  || ||  |  | || ||| | || |
Db    413 AAGCCGGCACCCCTGCAATTGTCGAGGCGATCGGCCTTGGTGCTGCGATCGACT 466


Search completed: April 22, 2003, 13:05:47
Job time : 1985.07 secs
```